Something old, something new:

The demanding process of serial updating in working memory

Shraddha Kaur

MRC Cognition and Brain Sciences Unit

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Christ's College September 2019

Dedicated to Chaiji, whom I miss very much,

and to the project that could never be.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other, university. This dissertation is the result of my own work, completed under the supervision of Prof Susan Gathercole and Dr Dennis Norris, and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. It does not exceed the prescribed word limit set by the School of Clinical Medicine and the Board of Graduate Studies.

Shraddha Kaur

September 2019

Something old, something new: The demanding process of serial updating in working memory

by Shraddha Kaur

Abstract

Working memory (WM) is a cognitive system that holds and manipulates information over short periods. The process of updating allows WM to selectively keep track of task-relevant information so that ongoing mental activity can proceed smoothly. The thesis established the first fine-grained temporal analysis of the cognitive demands associated with WM updating, while also presenting recall behaviour and self-reported strategy data. Chapter 1 provided a theoretical background and reviewed experimental investigations of WM updating with a particular emphasis on two updating tasks, running span and *n*-back. Chapter 2 then described an experiment in which the recruitment of cognitive resources was charted while participants completed a running span task. The chapter demonstrated that the resource demands associated with running span follow a specific time course that is absent in serial recall tasks that do not require WM updating. Chapter 3 examined whether the updating process profiled in the previous experiment is sensitive to the temporal parameters of the task. By tracking resource demands during running span with different rates of item presentation, the chapter showed that updating is time-consuming and thus limited to the slow-paced task. Chapter 4 presented a strategy induction procedure in which participants were instructed to perform running span using active updating or passive listening. This chapter compared the behavioural impact of the two strategies and replicated the time course of updating observed in the first two experiments using converging operations. Chapter 5 extended the investigation to another updating task and showed that the updating-related demands in *n*-back resembled those in running span. In addition, memory performance in terms of both serial position curves and recall/recognition errors was examined in each respective chapter. Self-report data were also presented in Chapters 3-5 to further understand the cognitive mechanisms of WM updating. Finally, Chapter 6 brought together findings from the four experiments and discussed them in the context of possible computational accounts of WM updating. It also considered the limitations of the present work and discussed possible avenues for further research.

Acknowledgements

This thesis would not have happened without the support of my family, supervisors, friends and colleagues, and wife-to-be. Thanks to…

my family, who gave me the gift of education and everlasting love. For their care and support, and their need to know if I am eating, drinking, sleeping enough. Ma, who first taught me how to formulate, test and revise hypotheses as we played Mastermind. Papa, who showed me that research takes more than an idea, it takes patience, perseverance and resolve. Pal, for his rainbow shoelaces and quiet cheerleading as I experimented with life outside research. And Shef, who is now basically family.

Sue, for being a fantastic supervisor and mentor. For always asking me "But what does that mean?" and pushing me to abstract away from task procedure and focus on cognition. For flexibly encouraging a change in topic, for helping me become a clearer thinker and better writer, and for believing in the project when I felt dispirited or sceptical.

Dennis, for humouring all the speculation about updating, for brainstorming about task design and appropriate controls, and for the suggestion of the modified span. For having a discerning eye that spotted writing mistakes despite multiple read-throughs.

the working memory team at the CBU for valuable discussions and keen feedback. In particular, Joni and Elizabeth, for support and understanding of challenges of updating (!), and great conference adventures in San Fran. Also, Rennie, for combing through and deciphering pages of strategy reports with me. Laura, for being the voice of my stimuli. Rogier, to whom I owe my R-skills. For his mentorship, and healthy doses of general assurance and anti-SPSS encouragement. Sally, for her stimuli, for her jolly company, and her wit and candour.

my fellow PhD students for our lovely board game evenings as we plodded along this academic journey; each separate and yet together in that office 59. In particular, Sneha and Hannes, for sharing triumphs and failures, delights and despairs, rum and coke.

the supportive, collaborative and intellectual environment at the CBU; the healing garden and the abundant tea and cake, which often served as my breakfast during those rushed days of participant testing. It is a grand feeling knowing it is in these very halls that much of cognitive psychology I cite today took place. Alan Baddeley, for his casual recommendation to use a self-paced design that became an entire study in its own right.

Gates-Cambridge Trust, Christ's College and the MRC for the financial support and community. My students, Sarbat, the LfP program, the GSP program, the German classes, Impact Labs and LinkAges, for giving me hard-to-refuse opportunities and keeping me busy outside of the PhD.

the 200+ participants who gave up their time and effort and collectively sat through 500+ hours of (incredibly dull!) tasks to bring us closer to understanding the mysteries of human memory.

and Julia, for inheriting me her desk and for inspiring the thesis title. For being the most wonderful companion in this PhDrama, and for showing me that it can be done (twice!). For encouraging (and sometimes mildly strong-arming) me to work when procrastination descended. For pulling a grumpy me into a kayak and reminding me that breaks are rejuvenating. For inviting me gently to put thoughts into words and waiting patiently when the words refused to flow. For being my human.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Overview

Working memory plays a crucial role in a wide range of cognitive activities by supporting the storage and processing of material over brief periods (Baddeley, 2012). It is a system with limited capacity and is unable to maintain a large amount of information indefinitely. There is thus a need for a mechanism that selects information relevant to our current goals and changes working memory contents as our goals change (Kessler & Meiran, 2008). This process is known as working memory updating, and it enables working memory to hold task-relevant information at any given time.

The focus of this thesis is on a particular kind of updating that occurs on the basis of serial order. For instance, air-traffic operators have to closely follow displays of information that keep changing with each new state of air traffic (Pollack, Johnson, & Knaff, 1959). The operators must rely on their short-term memory to recall the relevant information but often face uncertainty about *when* they might be required to report the latest display values. Considering this problem, Pollack et al. proposed that the operators would have to continually keep track of the presented material by effectively changing, or updating, the contents of short-term memory. Information presented earlier must be replaced by new information when capacity limitations are reached, or when the older information becomes outdated. To accommodate this, old (irrelevant) contents should thus be distinguished from new (relevant) ones in working memory. This demarcation should also be flexible so that information can be repositioned from the new to the old category once it becomes outdated. Previous work has widely researched how information is stored in working memory in the same order as it is presented (e.g., Hurlstone, Hitch, & Baddeley, 2014). How the old and new representations are separated and how the older representations are selectively discarded while the newer items are preserved remain outstanding questions. This thesis explores such serial-updating of working memory using different methodologies and in different paradigms. A consistent observation across four experiments is that updating is a demanding activity and is characterised by a distinct time-course generalisable across two tasks.

This introductory chapter presents the theoretical foundations of this thesis. It begins by introducing the influential models of working memory in Section 1.2. It then considers computational frameworks that simulate the short-term storage of incoming verbal information and how the serial order of encoded information is represented. This is followed by a discussion of how executive control in working memory is typically conceptualised and measured. Working memory updating is introduced in Section 1.3. This section describes a distinction between two categories of updating processes in working memory, item-updating and serial-updating. The computational and empirical evidence associated with serial-updating is then considered in more detail. Section 1.4 identifies behavioural methods that allow for the study of ongoing cognitive processes while participants engage in memory tasks and highlights the potential for these methods to provide deeper insight into the process of serial-updating. Finally, Section 1.5 outlines the overall aims and structure of the thesis.

1.2 Working memory

Working memory is a cognitive system that allows for the storage and processing of information over short periods (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Miyake & Shah, 1999). The concept of working memory was influenced by that of short-term memory proposed by Atkinson and Shiffrin (1968; for a review, see Baddeley, Hitch, & Allen, 2019). Working memory is considered a dynamic system that can store as well as process information, whereas short-term memory is thought to provide only passive, temporary storage (Aben, Stapert, & Blokland, 2012). It is integral to everyday functioning and involved in learning, academic performance, reading comprehension, following instructions, mental arithmetic, and reasoning (Adams & Hitch, 1997; Daneman & Carpenter, 1980; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008; Jaroslawska, Gathercole, Allen, & Holmes, 2016; Kane et al., 2004; Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Additionally, studies show that deficits in working memory functioning are closely associated with adverse developmental and ageing outcomes over the lifespan (Archibald & Gathercole, 2006; Babcock & Salthouse, 1990; Chiappe, Siegel, & Hasher, 2000; Cowan, 2014; Gathercole et al., 2016; Hasher & Zacks, 1988).

Several models of working memory have been proposed. The multiple components model of Baddeley and Hitch in 1974 is a highly influential model continuing to frame much of the research in the field. It proposes that working memory is composed of two specialised short-term storage systems, a domain-general executive control system and an episodic buffer that integrates information held in short-term and long-term storage. An alternative account proposed by Cowan conceptualises working memory as the activated portion of long-term memory with a few particularly salient chunks receiving maximum activation in the focus of attention (Cowan, 1999, 2005; Cowan et al., 2005). A third model proposed by

Engle, Unsworth and colleagues distinguishes two types of memory and suggests that attention can be allocated to either activate task-relevant representations or inhibit distracting information (Engle, Kane, & Tuholski, 1999; Unsworth & Engle, 2007; Unsworth & Spillers, 2010). A final model advanced by Barrouillet and Camos suggests that maintenance and processing activities in working memory are supported by the same general attentional resources and describes how these resources may be shared in time by rapid switching between the activities (Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009).

Although there is little consensus about the specific mechanisms that support working memory function, most models share two features. First, they view working memory as a system with limited capacity. An essential aspect of sustaining performance in a capacity-limited system is to exercise selective storage. Relevant information must somehow be selected while distracting, irrelevant information must be discarded or removed. This selective maintenance is a non-trivial process known as working memory updating and is the primary focus of this thesis. A second common property is an interplay between working memory and central attentional or executive control. It is this controlled interaction that facilitates the updating of working memory, although the proposed operations vary across the computational frameworks. The nature of capacity limitations and attentional control will be briefly described in the context of the multi-component model (Section 1.2.1.1), the embedded-processes model (Section 1.2.1.2), the time-based resource sharing model (Section 1.2.1.3) and the dual-component model (Section 1.2.1.3).

1.2.1 Theoretical frameworks of working memory

1.2.1.1 Multi-component model

Baddeley and Hitch proposed the multi-component model in 1974. It advanced three distinct components: a central executive and two supplementary domain-specific stores, the *phonological loop* and the *visuospatial sketchpad*. These temporarily store and maintain verbal and visuospatial information, respectively. A fourth component known as the episodic buffer was later added to the model to allow the integration of information across the various sub-components of working memory as well as to coordinate with long-term memory (Baddeley, 2000). See Figure 1.1 for an illustration of the multicomponent model.

The proposed fractionation of the working memory system has been tested using dual-task methodology. A central assumption of the model is that if two tasks that rely on the same component are performed simultaneously, they will interfere with each other and performance will be reduced compared with single-task conditions. Conversely, if two tasks engage different components, then performance should be insensitive to whether the tasks are performed together and separately. For example, it was found that continuously repeating the same word aloud (termed as articulatory suppression) impaired memory for verbal sequences while leaving the memory for spatial locations intact (Alloway, Kerr, & Langheinrich, 2010; Baddeley, Lewis, & Vallar, 1984). This outcome suggested that articulatory suppression taxes the phonological loop, and thus leads to verbal interference, but does not make use of the resources of the visuospatial sketchpad. Conversely, concurrent tapping of four corners of a square was associated with a reduction in spatial short-term memory but not verbal short-term memory, suggesting that it draws on the visuospatial sketchpad but not the phonological loop (Alloway et al., 2010; Vandierendonck, Kemps, Fastame, & Szmalec, 2004). Evidence from neuropsychological studies (for an overview, see Meiser & Klauer, 1999) and developmental studies (Alloway, Gathercole, & Pickering, 2006) provide further support for the domain-specific separation of processing streams.

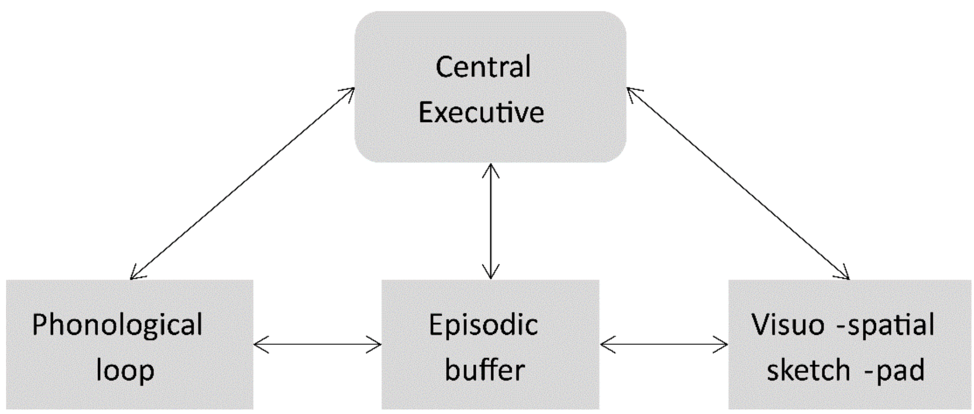

Figure 1.1 A simple representation of the multicomponent model, adapted from Baddeley (2000).

Phonological loop. The phonological loop is a specialised system to hold and process acoustic and speech-based information. It consists of two subcomponents, a phonological store that holds traces of verbal memory and an articulatory rehearsal mechanism that serves to revive the traces. The phonological store is often likened to the mind's ear. It receives sound-based input and holds it in the form of temporary phonological codes. These representations are passively held and thus decay after approximately two seconds (Baddeley, 1986). The traces need to be reactivated to ensure that they are protected from fading. This reactivation is done by the articulatory rehearsal mechanism, which involves covert or overt verbalisation of decaying phonological representations and can be thought of as the mind's voice (Baddeley, 2012). The phonological loop, therefore, transfers information in a loop, wherein it briefly enters the phonological store and is then refreshed by the articulatory rehearsal mechanism so that it re-enters the store and so forth.

Studies of the phonological similarity effect provide evidence for the phonological basis of representations in the phonological loop. The phonological similarity effect is the observation that stimuli that share acoustic features, such as the words *cat* and *rat*, are easily confused and thus harder to remember compared with stimuli that sound more distinct, such as the words *cat* and *dog* (Baddeley, 1966; Conrad, 1964; Conrad & Hull, 1964). Another key finding that supported the phonological loop was the word length effect. This effect is illustrated in higher memory performance for shorter words, e.g. *day*, *man*, *foot*, *rich*, compared with longer words, e.g. *identification*, *beautiful*, *conference*, *substitute* (e.g., Baddeley, Thomson, & Buchanan, 1975). This effect is observed even when the number of syllables is matched (Baddeley, 1986). In other words, a list of two-syllable words containing short vowels, and thus taking less time to verbalise, is recalled more accurately than a list of two-syllable words containing long vowels (e.g. *monkey* versus *cartoon*). The word length effect arises because the articulatory rehearsal mechanism involves covert or overt vocalisation in real-time. If articulation of the words takes longer than the rate of decay, some information is deteriorated and thus results in poorer memory than if the rate of decay is matched or exceeded by the rate of articulation (Baddeley, 1986, 2012).

It has also been suggested that the phonological loop evolved to support the acquisition of language (Baddeley, Gathercole, & Papagno, 1998). Its function is to retain the pattern or order in which unfamiliar sounds are arranged to form new words so that they can be remembered over the short-term while they are being consolidated into long-term memory. As such, the encoding of the *order* of incoming stimuli is obligatory and crucial for vocabulary acquisition. The major challenge confronted by computational models of verbal short-term memory, therefore, is to specify how item and order information is linked and retrieved. Section 1.2.2 provides a discussion of possible solutions.

Visuospatial sketchpad. The visuospatial sketchpad provides temporary storage and processing of both visual and spatial information (Baddeley, 1986). Since the original model was proposed, a wealth of empirical studies have shown that the sketchpad might be composed of separate systems for visual and spatial processing (Baddeley, 1996; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Logie, 1986, 1995; Logie, Zucco, & Baddeley, 1990). For instance, Della Sala et al. (1999) found that spatial working memory, measured using a Corsi blocks test, was disrupted in the presence of a spatial interference task that involved tracing a sequence of pegs on a wooden board. This spatial interference task did not impair performance in a test of visual working memory. The opposite pattern was found when a visual interference task that involved looking at abstract paintings resulted in a reduction in performance in the visual working memory test but not spatial working memory test. These data thus showed a double dissociation in the interference pattern, suggesting a separation between visual and spatial components of working memory. Neuropsychological and neuroimaging studies similarly suggest a modular dissociation in the brain regions engaged in visual and spatial memory (Baddeley, Bressi, Della Sala, Logie, & Spinnler, 1991; Carlesimo, Perri, Turriziani, Tomaiuolo, & Caltagirone, 2001; De Renzi & Nichelli, 1975; Luzzatti, Vecchi, Agazzi, Cesa-Bianchi, & Vergani, 1998; Shallice & Warrington, 1970; Vicari, Bellucci, & Carlesimo, 2003).

Logie (1995) advanced a fractionation of the visuospatial sketchpad by proposing two subcomponents. The first subcomponent is the *visual cache* that involves passive storage of visual features such as colour and form and is susceptible to decay. The second subcomponent is the *inner scribe* that provides a dynamic spatial process to rehearse and retrieve the contents stored in the visual cache. While it accounts for some neuropsychological data (Della Sala & Logie, 2002), the proposal does not accommodate separate mechanisms to maintain and rehearse visual and spatial information suggested by other studies (e.g. Klauer & Zhao, 2004). The precise mechanism serving spatial rehearsal is as yet unclear although it is suggested to involve covert oculomotor or eye-movement control (Pearson, 2006; Pearson, Ball, & Smith, 2014; Pearson & Sahraie, 2003). Pearson and colleagues showed that spatial short-term memory was impaired if a concurrent interference task prevented the preparation of covert eye-movements or required eye saccades. This finding is akin to the impairing effect of articulatory suppression in the phonological loop. Pearson (2006), however, proposed that the rehearsal of visuospatial material is more demanding than verbal rehearsal and might rely on domain-general resources in addition to modality-specific ones. Morey and colleagues put forward a more radical challenge to the concept of domain-specific spatial storage (Morey, 2018; Morey, Rhodes, & Cowan, 2019). They called into question existing experimental and neuropsychological evidence and argued instead in favour of moving away from modular, specialised systems. This position is currently the focus of vigorous debate (Logie, 2019).

Central executive. The central executive is responsible for coordinating and monitoring the domain-specific components to support complex cognition (Baddeley &

Hitch, 1974). In the original model, the authors considered the central executive as a limited pool of resources capable of general processing and attentional focus, along with some storage capacity and an ability to communicate with long-term memory (Baddeley, 1986; Baddeley & Hitch, 1974). Later versions of the theory adopted properties of the attentional control system proposed by Norman and Shallice (1986) and suggested further fractionation of the central executive (Baddeley, 1986; 1990; 1996). In this version, the central executive involved the capacity to focus, divide and switch attention, and the ability to provide an interface between the content in working memory and that in long-term memory. This final capacity was later transferred to the episodic buffer as described below (Baddeley, 2000, 2006). More recently, Logie (2016) challenged the central executive component, suggesting that executive control is an emergent property of interactions between specialist systems rather than a distinct system. He proposed that what appears a single, domain-general, flexible attentional capacity is supported by the seamless integration of information across specialised functions. However, he conceded that a singular entity might still provide a useful albeit simplified theoretical characterisation, especially when considering complex cognition. The concept of cognitive control in working memory and how it is measured is further considered in Section 1.2.3.

Episodic buffer. The episodic buffer was added to working memory by Baddeley in 2000. Its purpose was to support the integration of information between the phonological loop and visuospatial sketchpad in a multidimensional code. These multidimensional representations were said to be held as episodes within working memory for brief periods, with both perception and long-term memory. The capacity of the store was approximately four integrated chunks or episodes (Baddeley, 2003). Recent evidence suggested that the episodic buffer is a passive store. While it could hold integrated episodes for conscious access, the buffer itself does not appear responsible for the integration or binding of an episode (Baddeley, Allen, & Hitch, 2010). Baddeley argued that retrieval from the buffer requires conscious awareness and is coordinated by the central executive (Baddeley, 2000, 2003, 2011).

1.2.1.2 Embedded-processes model

Cowan advanced an alternative framework of working memory known as the embedded-processes model (Cowan, 1999, 2001; Cowan et al., 2005; Cowan & Rachev, 2018). This model characterises working memory as the activated and thus easily accessible portion of long-term memory, rather than treating long-term and short-term memory as distinct stores and processing systems. See Figure 1.2a for an illustration of the model.

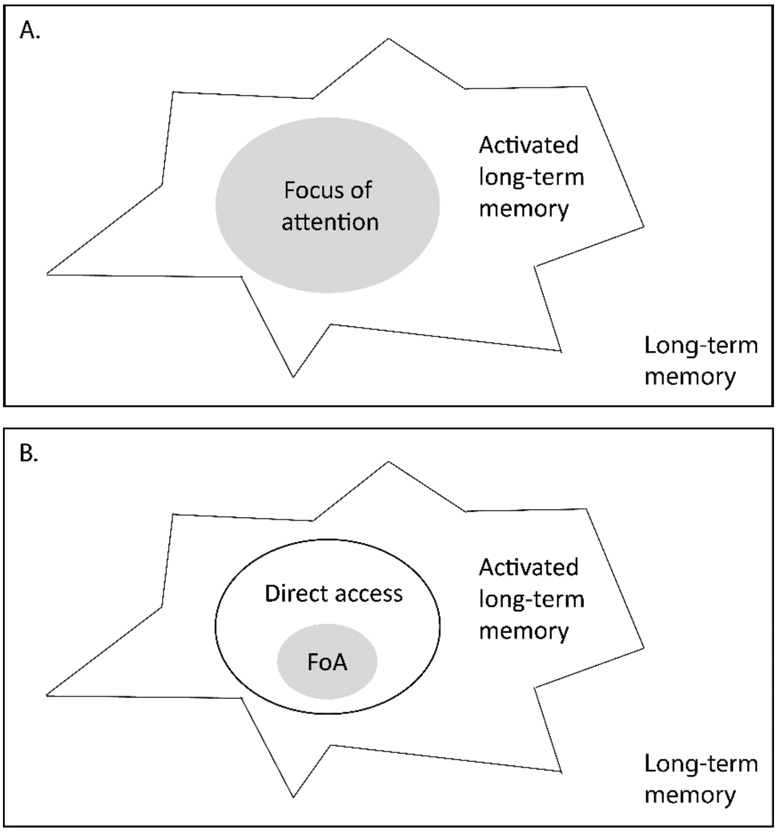

Figure 1.2 A simple representation of (a) Embedded processes model proposed by Cowan (1999), and (b) Concentric regions model proposed by Oberauer (2002).

According to Cowan and colleagues, contents in long-term memory are stored in one of three states: (i) (un-activated) long-term memory, (ii) activated long-term memory (working memory), and (iii) focus of attention within working memory (Cowan, 1995, 2008). The majority of the information in long-term memory is held with relatively low levels of activation as it is not relevant to the current task. A small portion of this information receives an activation boost as a result of ongoing mental activity and constitutes working memory. Task-relevant information is maintained in activated long-term memory in a readily accessible state for brief periods. There is a smaller part of activated long-term memory, termed as the focus of attention, in which a limited set of salient representations are held. Typical estimates suggest that the focus of attention can simultaneously hold around four chunks of features bound together meaningfully (Cowan, 2001; Cowan et al., 2005; Gilchrist & Cowan, 2011).

The model also incorporates a domain-general attentional component composed of an involuntary and a voluntary system (Cowan, 1999). The involuntary system brings information into the focus of attention on the basis of stimulus salience, while the voluntary system provides top-down executive control. Such executive control is key to achieving and maintaining task-relevant information in the focus of attention available so that it is readily available for conscious access. Information related to an unchanging external environment or no longer relevant to the current task is not actively maintained and decays over time. Several operations have been suggested to reactivate decaying representations, including subvocal rehearsal, mental imagery and rapid scanning to circulate the items through the focus of attention (Cowan, 1992, 1995).

Oberauer in 2002 reviewed empirical evidence for the role of the focus of attention. He found that some studies suggested a broad focus in line with Cowan's characterisation capable of holding four chunks at any given time, while others favoured a more restricted interpretation suggesting it could only hold a single chunk. On this basis, Oberauer put forward a concentric regions model with three separate regions of working memory: activated long-term memory, a region of direct access as well as a focus of attention (2002, 2009, 2013). This model is illustrated in Figure 1.2b. The region of direct access retains a few items from activated long-term memory in a prioritised state to support current cognitive activity. It also allows the binding of disjointed stimulus features and context representations into a single chunk. The focus of attention holds a single item or chunk that is relevant to the next operation required to achieve the current goal. This model proposes that activation does not decay over time. Instead, capacity constraints arise due to limitations in central attention or increased interference between features of different working memory representations. When faced with a memory test, participants might experience an intrusion (i.e. incorrect retrieval) of non-target information that was previously task-relevant as it may have some residual activation due to its previous relevance to the task. This residual activation is said to generate a preliminary "familiarity" signal. A more reliable, but slower "recollection" signal is obtained from the bound information in the region of direct access. Retrieval of this information allows identification of not only an item itself but also the context features of the item such as serial position in the list. These signals will be revisited in a later section when discussing interference effects in the *n*-back task (see Section 1.3.4).

1.2.1.3 Time-based resource sharing model

The time-based resource sharing model (TBRS) of working memory was developed by Barrouillet, Camos and colleagues in 2004. This model, illustrated in Figure 1.3, describes the attentional processes rather than the structure of the working memory system (Barrouillet et al., 2004, 2009).

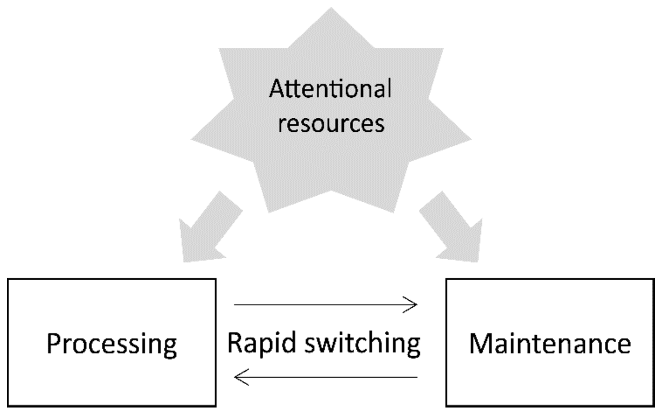

Figure 1.3 A simple representation of the Time-based resource sharing model of proposed by Barrouillet et al. (2004).

According to this model, attention supports both the processing and maintenance of information in working memory. Items maintained in working memory undergo decay as soon as attention is allocated away from them. A process of attentional refreshing is thus required to retrieve the decaying representations from memory into the focus of attention and serve current goals. This is done by refocusing attention away from the processing activity and toward the items stored in working memory. Therefore, the maintenance and processing activities of working memory share limited attentional resources. Crucially, this model proposes that the resource sharing between activities is time-based (Barrouillet et al., 2004; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet & Camos, 2012). This is because it involves memory retrievals of decaying items, which are subject to a central bottleneck and can only occur one at a time. As a result, the sharing of resources is accomplished by frequently and rapidly switching attention between ongoing processing and maintenance demands. Such rapid switching between the two sources of demands in working memory tasks was not a novel proposal. For instance, it was previously suggested that switching between maintenance and processing could facilitate the intermittent *rehearsal* of the stored information and protect it from complete loss from working memory (Towse, Hitch, & Hutton, 1998). However, the TBRS model argues that activities such as rehearsal, while still possible, are time-consuming. It thus proposes that item reactivation via attentional refreshing provides a more effective alternative mechanism (Barrouillet et al., 2004; Camos & Barrouillet, 2014).

1.2.1.4 Dual-component model

Engle and colleagues proposed the dual-component model of working memory (Engle & Kane, 2004; Engle et al., 1999; Unsworth & Engle, 2007; Unsworth & Spillers, 2010). This framework distinguishes two memory systems: Primary and Secondary memory (Figure 1.4, adapted from Rose & Craik, 2012). Primary memory can hold around four distinct representations or chunks of information relevant to current cognitive processing. Attentional resources are required to actively maintain the contents in Primary memory and protect them by suppressing or inhibiting distracting information. Secondary memory contains information not actively maintained at any given point. This information could have been transferred from Primary memory to Secondary memory due to limitations in capacity or a failure in the active maintenance systems. Retrieval from Secondary memory into Primary memory relies on a probabilistic search mechanism based on cues (e.g. temporal, categorical, or contextual). The use of effective cues delimits the search set and enables successful retrieval. According to this model, individual differences in capacity arise due to variation in attentional control in Primary memory as well as retrieval efficiency in Secondary memory (Engle et al., 1999; Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001).

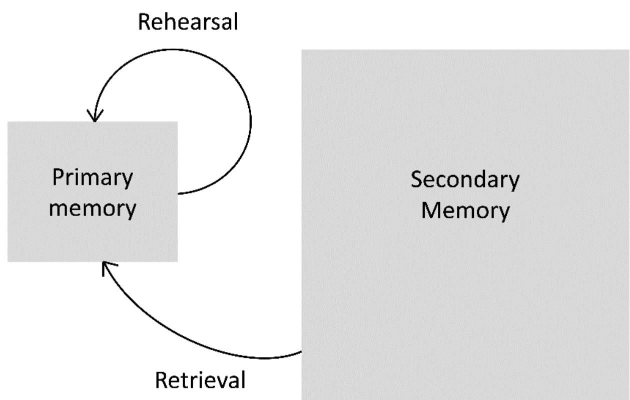

Figure 1.4 A simple representation of the dual-component model proposed by Engle, Unsworth and colleagues, adapted from Rose and Craik (2012).

In summary, the above section reviewed the main models of working memory, namely the multi-component model, the embedded-processes model, the time-based resource sharing model and the dual-component model. The structure, processes and functions of working memory continue to drive robust debate in the field (Cowan, 2019; Logie, 2016, 2019; Morey, 2018; Morey et al., 2019; Norris, 2017, 2019). Despite contrasting arguments, it is clear that across all models, capacity limitations are applied to the information stored in an accessible state to serve behaviour and thought. Another common feature is that all models propose that central executive or attentional resources are essential in manipulating, and sometimes also maintaining, the contents of working memory. The next section discusses a few prominent computational models that specify how item and order information is stored in the short-term as well as highlighting how the information is lost or removed from storage.

1.2.2 Computational models of serial recall

Serial order is fundamental to everyday activities (Lashley, 1951). Sequencing and processing information in serial order is necessary for a range of verbal and non-verbal activities such as learning new vocabulary and spelling, perceiving speech, implementing motor control and planning goal-directed action (Hurlstone et al., 2014). A number of computational models simulating verbal serial recall have been proposed over the last three decades (Brown, Neath, & Chater, 2007; Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1992, 1999, 2006; Farrell & Lewandowsky, 2002; Henson, 1998; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Page & Norris, 1998, 2009). Here, three models are described below as these will be particularly relevant when considering the process of updating based on serial order later. See Figure 1.5 for an illustration of these models. It is important to note that only the models of verbal serial recall are considered here; for a detailed consideration of computational principles underpinning visuospatial short-term memory, see a review by Hurlstone et al. (2014).

1.2.2.1 Network model of the phonological loop.

Burgess and Hitch in 1992 described verbal short-term memory as a two-layer feed-forward neural network, in which a layer of stored items was connected by a set of weights to another layer representing contextual factors such as serial position. During input, the phonemic information of the presented item was represented in the item layer, and the current position was represented in the context layer using an event signal (Burgess & Hitch, 1992) or a temporal coding scheme (Burgess & Hitch, 1999). Item-position associations were said to be learnt using a one-shot Hebbian adjustment of the weight matrix between the two layers. These associations were assumed to decay with time and required active rehearsal to be maintained (simulated in the model as covert recall).

The network model proposed by Burgess and Hitch (1992, 1999) accomplished recall in the following manner. Output was initiated by reinstating the context marker (event or timing signal) and moving it along each successive position representation. Competitive queuing was used to select the associated item among noisy activation values. This item was reported in the form of output phonemes for (covert or overt) recall followed by inhibition of the emitted item. During rehearsal, the covert output was re-presented as input, creating a phonemic feedback loop that reactivated the decaying item-position associations. It was not specified how the model reinstates the marker or context code for the first position when faced with upcoming retrieval. It was simply assumed that this would be the case to enable effective rehearsal and recall.

1.2.2.2 Primacy model

The primacy model, developed by Page and Norris, invoked a representation of serial order in terms of activation strength of list items (1998; for a revised model, Page & Norris, 2009). In this model, the first item in a sequence was the most strongly activated and the magnitude of activation decreased for every successive item. Item order was thus directly encoded in the activation gradient of the items, without an independent or external context marker to designate the position in the list. Recall was driven by competitive queuing on the basis of activation levels combined with the process of response suppression. The item with the largest activation at any given point was selected for output via a noisy-choice process, and the chosen item was then suppressed to allow the retrieval of the item with the second-largest activation. This suppression prevented a repeated recall of the same item over multiple retrieval attempts. As in the model discussed above, time-based decay and cumulative rehearsal were also fundamental to the primacy model. Active rehearsal was applied to protect against the decay and operated as a re-presentation of the entire encoded sequence to maintain relatively robust activation levels. An important distinction, among others, between the primacy model by Page and Norris and the network model by Burgess and Hitch was the lack of reliance on an independent context to drive output at recall. While Burgess and Hitch guided the position code to the first position at recall, Page and Norris relied on the activation strength to implicitly determine the first item to be recalled.

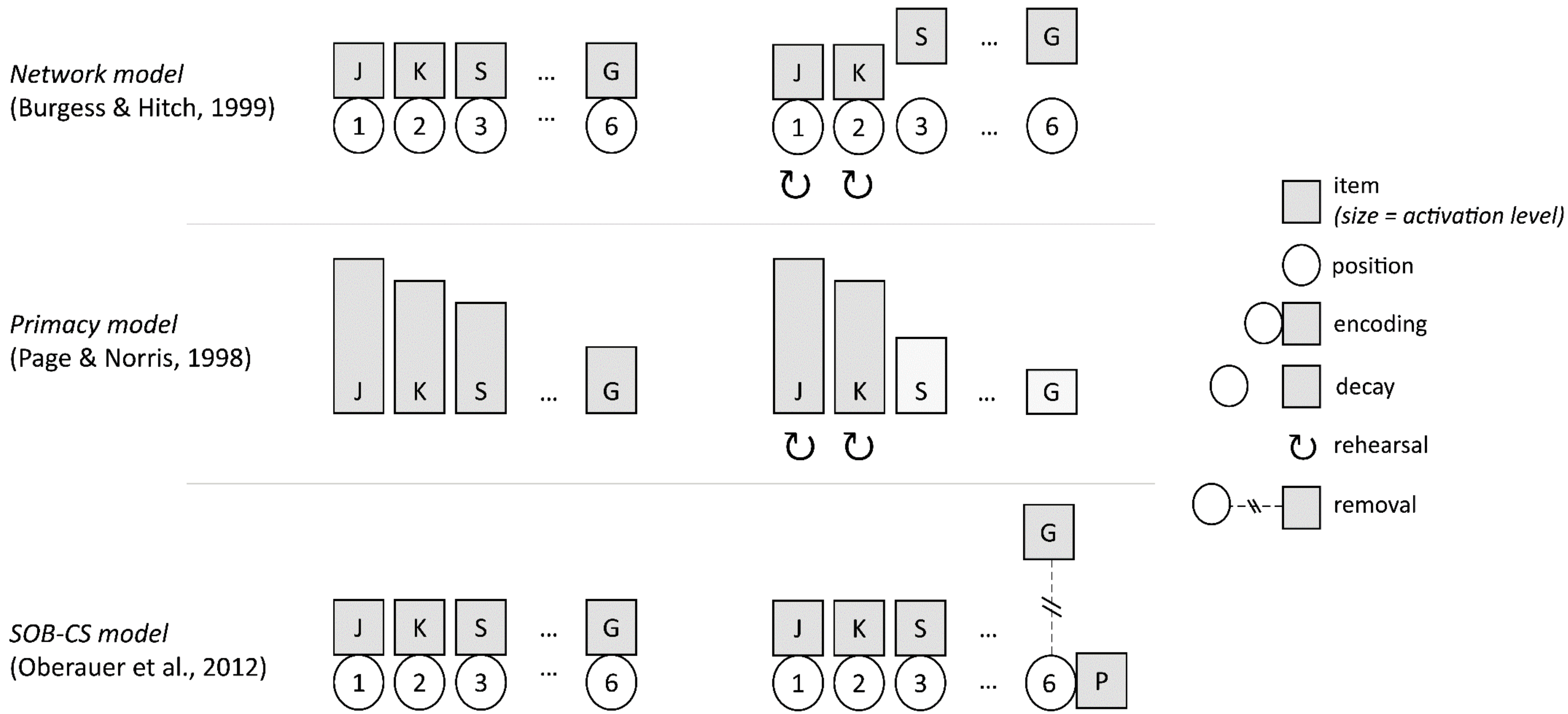

Figure 1.5 An abstract schematic of three prominent computational models of serial recall (a) the network model of the phonological loop proposed by Burgess and Hitch (1999; *top*), (b) the Primacy model proposed by Page and Norris (1998; *middle*), and (c) the SOB-CS model proposed by Oberauer et al., 2012; *bottom*).

1.2.2.3 Serial-order-in-a-box

Farrell and Lewandowsky in 2002 introduced a distributed neural network termed serial-order-in-a-box (SOB). During encoding, the items were auto-associated, i.e. every item was associated with itself, using Hebbian learning with a single presentation. The encoding strength varied such that successive list items were associated with progressively less energy. Such variation in energy dynamics naturally generated a primacy gradient in a similar manner as in the Primacy model described above. This was later modified in a second version termed C-SOB in which a two-layer structure was introduced. The prefix was "C" was added to the SOB model to note the explicit representation of context, i.e. position markers instead of a reliance on an emergent activation gradient (Farrell, 2006; Lewandowsky & Farrell, 2008). In this model, encoding occurred through Hebbian associations between item and position representation and capacity limitations were said to originate from interference between item-position associations. The next iteration of the model called SOB-CS ("CS" denoted complex span) provided a mechanism to process distractor information but minimise the interference by clearing irrelevant content from working memory (Oberauer et al., 2012). Distractor removal was modelled simply as a reversal of the encoding process. Hebbian antilearning was said to proceed gradually to unbind the irrelevant item from its associated position (pp.787-8).

To summarise, the three models described above provide different solutions to implement serial recall of recently encoded information. They differ in their positional representations, preserving the serial order of a sequence by either employing an absolute or relative coding (Burgess & Hitch, 1992, 1999; Farrell, 2006; Oberauer et al., 2012) or invoking activation gradients (Farrell & Lewandowsky, 2002; Page & Norris, 1998, 2009). The models also attribute the capacity limitations of short-term verbal memory to different sources. In some models, working memory is considered to be a system vulnerable to rapid decay (Burgess & Hitch, 1992, 1999; Page & Norris, 1998, 2009), whereas the SOB model and its derivatives consider limitations in short-term storage arising from increased interference between representations (Farrell & Lewandowsky, 2002; Oberauer et al., 2012). Accordingly, the operation that is engaged to overcome the capacity-limitations is different. While the decay-based models posit an active process of rehearsal to reactivate representations, the interference-based models incorporate the reverse logic. In these models, all content is maintained without decay, and active removal is required to reduce the influence of distracting or interfering information. These alternatives are very much the matter of recent debate as researchers consider how forgetting, or loss of information, occurs in short-term memory (Barrouillet et al., 2007; Barrouillet, Portrat, Vergauwe, Diependaele, & Camos, 2011; Farrell et al., 2016; Lewandowsky & Oberauer, 2009, 2015; Oberauer & Lewandowsky, 2008, 2014; Portrat, Barrouillet, & Camos, 2008). Since working memory updating in its simplest form requires the *forgetting* of outdated information and the maintenance of relevant material, the decay versus interference debate will be reconsidered when discussing models of working memory updating in Section 1.3.2.

1.2.3 Executive control in working memory

The theoretical and computational models of working memory discussed in the previous sections are all capable of simulating experimental phenomena related to verbal serial recall. Some models view working memory as a system composed of both domain-specific and domain-general components (e.g. Baddeley & Hitch, 1974), while others place more emphasis on a domain-general structure (Cowan, 1999; Engle et al., 1999; Oberauer, 2013). However, all models assume cognitive control in some form to support complex cognition, either as a central executive component within working memory or an external component that controls attentional resources. The exact nature of these components or resources remains underspecified, with researchers often using terms such as executive attention to refer to both the attention directed to goals and actions and the control of attention. Oberauer (2019) proposed a recent taxonomy of attention and reviewed how it relates with the working memory system. This thesis primarily treats attention as a limited resource serving control processes and uses the more neutral term of general cognitive resources.

A thorough understanding of the mechanisms underlying cognitive control required during everyday tasks is clearly limited. However, there have been advances in understanding how cognitive control is applied in specific working memory paradigms. These share a common feature in that participants must undertake some form of additional processing beyond simple storage and maintenance of the items to be recalled. The specific requirements can vary widely across the paradigms, in some cases even imposing demands and constraints that often do not resemble familiar, everyday situations. These requirements, as well as proposed cognitive solutions, will be briefly considered below in three exemplar categories of tasks that are frequently used to examine cognitive control.

Complex span. The first category of tasks, called complex span tasks, is commonly used to examine the storage and processing components of working memory (Daneman & Carpenter, 1980; Engle, Laughlin, Tuholski, & Conway, 1999; Kane et al., 2004; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009). In these tasks, the intervals between successive memory items (e.g. letters) are filled with episodes of a processing task (e.g. symmetry judgements of shapes). The distractor processing activity is usually unrelated to the to-be-recalled items, and the load on working memory can be manipulated by varying either the number of items to remember or the number of processing episodes to complete between two successive items. The central requirement in these tasks is for participants to protect the memory items from either decay or interference caused by the interpolated distracting activity. This might be achieved in various ways. Participants may switch rapidly between distractor processing and memory maintenance by undertaking either rehearsal (Towse et al., 1998), or attentional refreshing of memory items (Barrouillet et al., 2009). Such time-switching enables the recovery of stored representations that are thought to be decaying as participants complete the interpolated distractor activity. Alternatively, encoded memory items may be shielded from the interference created due to processing of the distractor items using a gating mechanism (Oberauer et al., 2012). By this account, the opening and closing of this proposed gate determines whether perceived information is encoded into working memory, limiting interference.

Resequencing tasks. The second category of complex working memory tasks requires participants to re-sequence memory items so that they are recalled in an order different from the one in which they were presented. One example of such resequencing is a task in which mixed lists containing both numbers and letters are presented, and participants are asked to recall the numbers in numerical order and letters in alphabetic order (Sheslow & Adams, 1990; Weschler, 1997). Another common task is backward span in which participants have to report the encoded sequence in reverse order. It has been long assumed that participants accomplish this by undertaking a forward retrieval of the list, peeling off the final item, truncating that item from the list, and then repeating the process until the entire sequence is recalled, for example if the presented sequence is A,B,C,D, then response using peel-off would be the following: (A,B,C,)D … (A,B,)C … (A,)B … A (Anders & Lillyquist, 1971; Conrad, 1965; Thomas, Milner, & Haberlandt, 2003). More recent response time analyses indicate that such a peel-off strategy is rarely employed and that participants employ a variety of other strategies to perform the task (Norris, Hall, & Gathercole, 2019).

Updating tasks. The third category of tasks involves memory updating or keeping track of a changing set of memory items. In these tasks, information that was relevant at one time-point may no longer be relevant at another time-point during the task. The older, outdated information needs to be updated to keep only the relevant material in an accessible state. Working memory updating and its measurement will be discussed in Section 1.3. Two sub-classes of updating tasks requiring item- and serial-updating will first be contrasted. Potential computational implementations of serial-updating will then be outlined, followed by a review of the empirical investigations of the running span and *n*-back tasks.

1.3 Working memory updating

The theoretical frameworks and computational models discussed above describe how (ordered) information may be encoded, stored and maintained in working memory over short periods of time for retrieval. However, they do not describe the ways in which a system can be readily configured to deal with changing targets or goals in response to changing environmental or *ad hoc* task demands. Consider the complex task of multiplying two two-digit numbers. Such a problem would likely require multiple computational steps with interim results that would, in turn, be combined to reach a final outcome (Fürst & Hitch, 2000). This would, therefore, require the working memory system to temporarily protect the interim results from decay and/or interference but then modify them for the final step. This poses concurrent demands on working memory to maintain stable yet flexible representations, allowing us to hold information relevant to any given activity while letting go of (and not being distracted by) the irrelevant information. This is referred to as the ability to update working memory (Bunting, Cowan, & Saults, 2006; Ecker, Lewandowsky, & Oberauer, 2014; Ecker, Oberauer, & Lewandowsky, 2014; Kessler & Meiran, 2008; Morris & Jones, 1990).

Updating is also one of the three executive functions proposed by Miyake and colleagues (2000). Importantly, updating appears to differ from the other executive functions such as switching and inhibition as it is the only one that predicts individual differences in fluid intelligence (Chen & Li, 2007; Friedman et al., 2006). Working memory updating is also vital to many everyday mental activities including arithmetic calculations (Deschuyteneer, Vandierendonck, & Muyllaert, 2006), navigating (Garavan, 1998; Gugerty, 1997), reasoning (Carpenter, Just, & Shell, 1990), reading comprehension (Carretti, Cornoldi, De Beni, & Romano, 2005; Palladino, Cornoldi, De Beni, & Pazzaglia, 2001) and problem-solving (Cornoldi, Drusi, Tencati, Giofrè, & Mirandola, 2012; Passolunghi & Pazzaglia, 2005). These tasks all require a cognitive apparatus to attend to an ongoing, rapidly changing stream of information and continuously extract relevant information in order to accomplish our goals. In fact, it is argued that working memory updating is crucial in maintaining the "correct worldview" in the face of misinformation and misconceptions (Kessler & Meiran, 2008, p. 1339; Lewandowsky & Heit, 2006).

1.3.1 Categories of updating

Laboratory tasks that have been used to study working memory updating include alpha span, keep track, reference-back, running span, *n*-back (Kane et al., 2004; Morris & Jones, 1990; Rac-Lubashevsky & Kessler, 2016; Schmiedek, Li, & Lindenberger, 2009). These tasks all have in common the need to modify, replace, or recode the encoded information. The precise demands vary considerably across tasks. Kessler and Meiran (2008) suggested that several cognitive processes may be used to accomplish working memory updating and that a different process or set of processes may be required depending on the updating context. Here, two broad updating categories are proposed on the basis of the specific component of memory that needs to be updated: item-updating and serial-updating. Table 1 provides exemplars for each proposed category, and Figure 1.6 illustrates a trial from one task typical of each category.

Table 1.1

An overview of the two categories of updating with exemplar tasks for each category.

Category	*Example*	*Description of task requirements*
Item updating	Keep-track	Replace items in one of *n* semantic categories at every step, and recall the latest item in each category (Friedman et al., 2006; Yntema, 1963)
Item updating	Memory updating	Replace or manipulate items in *n* spatial locations, and recall the latest item in each frame (Ecker, Oberauer, et al., 2014; Kessler & Meiran, 2008)
Serial updating	Running span	Serial recall of the latest *n* items from a list of unknown length (Bunting et al., 2006; Morris & Jones, 1990)
Serial updating	*N*-back	Compare the current item with the one presented *n* positions ago (Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Kirchner, 1958)

In item-updating, each presentation cycle represents a local change in one (or more) of the encoded set of items, say through transformation or replacement, and the memory of the final state of each item is tested after a set number of update cycles. Experimental investigations of item-updating and associated theoretical models are discussed in Section 1.3.1.1. In serial-updating, only the latest information in a running series of items is relevant at any given time, but it is unclear when in the series memory will be tested. So, the items themselves are not updated, but whether or not they are relevant for retrieval changes as the presented sequence grows longer. Serial-updating is considered further in Section 1.3.1.2, followed by a discussion of possible theoretical mechanisms and experimental investigations.

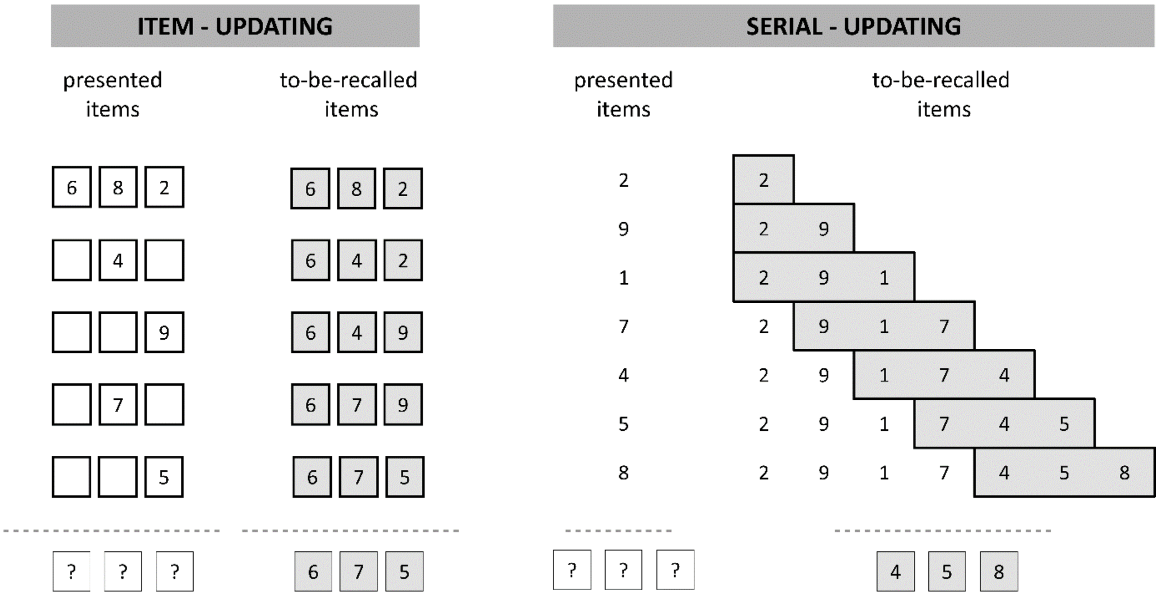

Figure 1.6 An illustration of a representative trial in a generic item-updating paradigm (*left*) and serial-updating paradigm (*right*). Each row represents one cycle of updating and the last row represents the final recall set. The grey shaded areas represent the target set to be recalled at any point in the task.

1.3.1.1 Item-updating

Item-updating is the process of implementing changes in the stored information at the item level. One of the earliest tasks involving item-updating is the *keep track* task devised by Yntema and Mueser in 1962 (see also, Friedman et al., 2006; Yntema, 1963). In this task, a number of target categories, such as animals, sports, colours, are presented, and participants have to parse an incoming stream of stimuli, such as words or images, into these categories. The objective is to keep track of the last exemplar of each category for eventual recall.

Another example is the *memory updating* task used by Oberauer and colleagues (e.g., Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014; Kessler & Oberauer, 2015). These involve updating by changing internal representations, for instance converting '8' to '7' if the task requires a subtraction of each cued item by 1, or updating the representation of an arrow from pointing upward to rightward if the task requires a 90° rotation.

Cumulative findings from theoretical, experimental and latent variable analyses have suggested that item-updating may be composed of three component processes: retrieval, transformation and substitution (Ecker, Lewandowsky, et al., 2014; Ecker, Lewandowsky, Oberauer, & Chee, 2010a; Ecker, Oberauer, et al., 2014). The studies further indicated that the process of substitution might be divided into two sub-processes: the removal of old, irrelevant information and the encoding of new, relevant information. In their 2014 study, Ecker and colleagues presented items in one of three frames, and this item replaced the previous one in that frame. The components of encoding and removal were disentangled using two experimental manipulations. First, the to-be-updated frame was cued before the replacement item was presented to alert participants to the specific item that needed changing. Second, the interval between the update cue and the presentation of the new item was varied. It was found that participants took advantage of long cue-target intervals to engage in item removal even before the presentation of the replacement item. Thus, response latencies after target presentation during long cue-target intervals were faster as they only reflected the time required for encoding. In the case of short cue-target intervals, response latencies were slower as they included both the time taken to remove the old item and encode the new item. This demonstrated that removal was separate from encoding, and the two processes could occur independently when appropriate cues were provided.

On this basis, the scanning and gate switching model was developed (Kessler, 2017; Kessler & Oberauer, 2014, 2015). This model considers item-updating an active process that requires a gated switch from a maintenance mode to an updating mode. Kessler and Oberauer (2015) showed that participants scan the encoded sequence in forward order, at least when the sequence is systematically ordered, e.g. reading a visually presented sequence of English letters from left to right or Hebrew letters from right to left. According to this model, a mode switch is triggered as soon as the to-be-updated item is encountered. The old item is then removed using a Hebbian anti-learning process to unbind the item representation from its positional code, extending the serial-order-in-a-box or SOB model as described earlier (Oberauer et al., 2012). Item removal is followed by the opening of the gate so that new perceptual input would be encoded into working memory using Hebbian learning, binding it to the same position from which the old item was removed. It was found that updating latencies capture the three processes of item removal, mode switching and item encoding, although it is the removal process that appears to be specific to updating (Ecker, Oberauer, et al., 2014). A recent study also found that the efficiency of item removal was related to both working memory capacity and fluid intelligence (Singh, Gignac, Brydges, & Ecker, 2018). Mediation analysis showed that the relationship between removal efficiency and fluid intelligence was entirely mediated by the influence of removal on working memory capacity. The authors suggested that updating contributes to reasoning ability by reducing the interference in working memory, thereby increasing memory capacity.

Removal by unbinding can be distinguished from a temporary, deactivation process. The latter is a process that leaves the item-position bindings intact but inactivates the item representations for a short duration (Lewis-Peacock, Kessler, & Oberauer, 2018; Oberauer, 2018). The inactive items thus do not compete for retrieval temporarily but continue to be stored in memory and can be retrieved by refocusing attentional resources on their bindings. Further, several behavioural investigations studying updating latencies found that the process of updating an entire memory set was faster than the process of updating it partially, say one item at a time (e.g. Ecker, Oberauer, et al., 2014; Kessler & Meiran, 2008). This distinction between complete, wholesale and partial, item-wise updating has been supported by neuroimaging investigations (Murty et al., 2011) and was simulated computationally by rapidly resetting the weight matrix between item and position layers (Ecker, Oberauer, et al., 2014; Oberauer et al., 2012). Therefore, updating latencies associated with a complete memory reset only reflect encoding time, as eliminating the entire contents of working memory is a prompt process. The time costs associated with removal as described above are only incurred in the case of partial, item-wise updating, in which some items are removed from working memory while others continue to be maintained.

1.3.1.2 Serial-updating

The objective in serial-updating tasks is to maintain the latest *n* items in an ordered manner from a list of unknown length. Here, updating can be thought of as a moving rehearsal window containing the target retrieval set of *n* items (Juvina & Taatgen, 2007). With every new presentation, the encoded list lengthens, and the rehearsal window shifts by one position. The items do not undergo any change per se and also continue to retain their relative serial order. What changes is the absolute position of each item in the rehearsal window. With a few updating steps, a recently encoded item might shift through the length of the rehearsal window one position at a time until it eventually moves out of the window and is then no longer relevant for recall. This is illustrated in Figure 1.6b, in which the fourth encoded item in the serial updating trial, *7*, is encoded at the final target position and successively moves across the length of the target window until it is not in the target set in the final updating cycle.

Updating on the basis of serial order therefore involves a change in the representation of both the irrelevant and relevant items, unlike item-updating in which only the to-be-updated item changes. In serial-updating, the irrelevant item must be removed, and the remaining (relevant) items must change their positions so that the new item can be incorporated into the target set.

Two tasks, running span and *n*-back, involve serial-updating. A typical running span task proceeds by presenting lengthy stimulus sequences and testing memory using serial recall of the latest set of *n* items (Bunting et al., 2006; Morris & Jones, 1990). Other versions of running span test recall or recognition memory using probes rather than complete serial recall of *n* items but the presentation procedure remains identical to the typical version of the task (e.g., Ruiz & Elosua, 2013). The second serial updating task is *n*-back, in which memory is tested using recognition decisions *during* the sequence rather than serial recall at the end. As new items are presented, judgements are required whether the latest item is the same as the one presented *n* positions back in the sequence (e.g., Jaeggi et al., 2010; Kane, Conway, Miura, & Colflesh, 2007).

Notably, the cognitive demands of *n*-back and running span are not identical. While running span simply involves item encoding and updating during the presentation phase, the *n*-back task involves an additional, concurrent matching and decision process. The tasks, at least in the standard versions, also vary in their retrieval demands. Running span requires serial recall of *n* items while *n*-back demands single item recognition. Despite these differences, it has been argued that *n*-back and running span are cognate procedures recruiting the same process of serial-updating (Ruiz & Elosua, 2013). In support of this account, a training study shows cross-task benefits. Performance in *n*-back improved as a result of intensive, adaptive training with running span tasks (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008; but see Bäckman et al., 2017; Zhao, Xu, Fu, & Maes, 2018). Further, in neuroimaging studies, running span and *n*-back tasks were both associated with striatal activity (functional magnetic resonance imaging) and striatal dopaminergic activity (positron emission tomography). Together, the two studies suggest an overlap between the two tasks, arguably in terms of the serial-updating process. However, the cognitive mechanisms of this process remain unclear. Section 1.3.2 will consider alternative theoretical accounts, derived from models of serial order previously discussed.

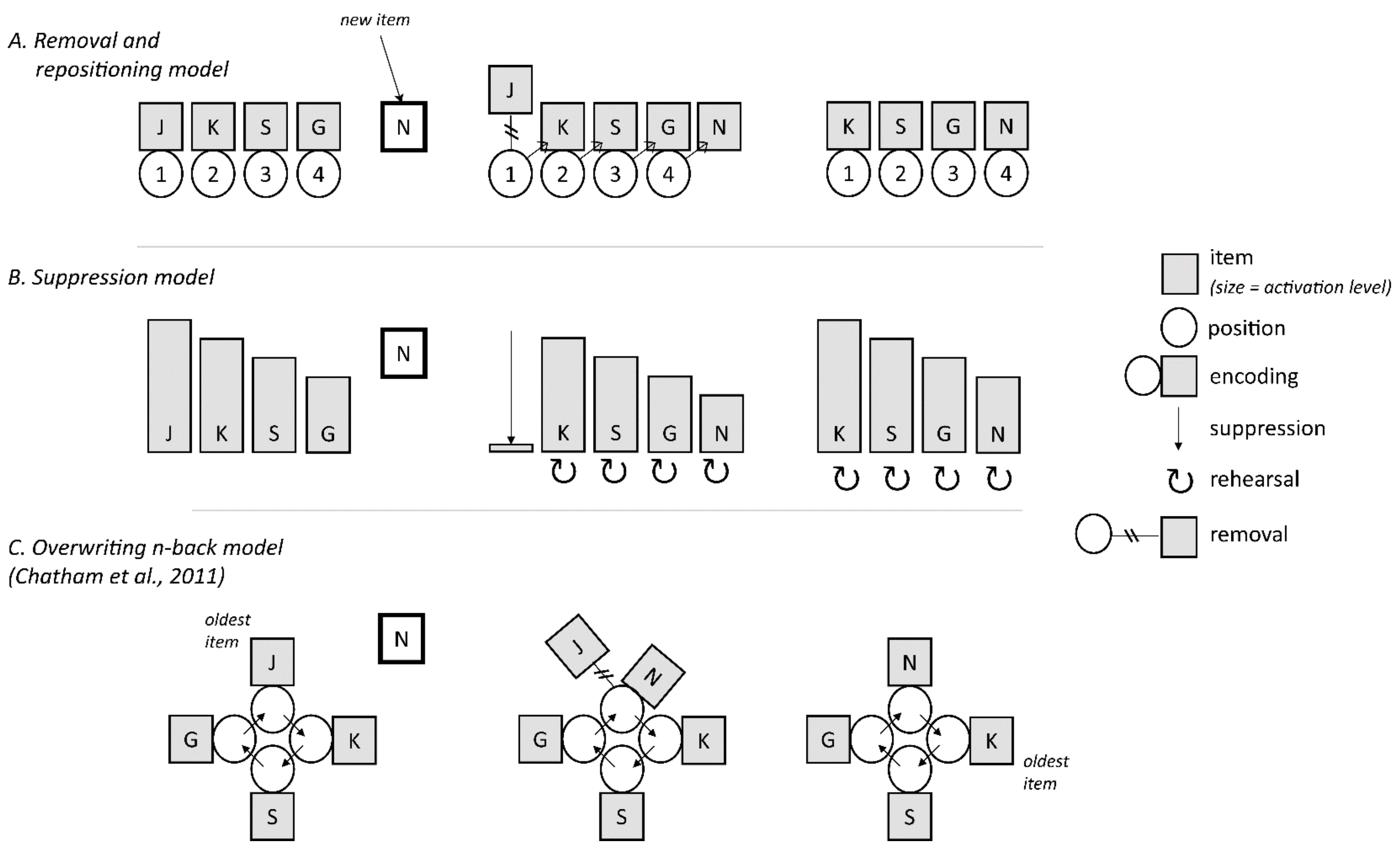

Figure 1.7 An abstract schematic of three models that could accomplish serial-updating. (a) the removal and repositioning model, adopting an item-removal mechanism by Oberauer et al. (2012). (b) the suppression model, adopting a primacy gradient as proposed by Page and Norris (1998). (c) the overwriting *n*-back model proposed by Chatham et al. (2011), with a cyclical representation of serial order. See text for more details.

1.3.2 Modelling serial-updating

A drop-and-capture framework of updating was advanced by Pollack et al. (1959) when they devised the running span task. It was proposed that the short-term memory would hold the target set comprising the last few items of a presented sequence, pre-determined based on the task or individual capacity. After the presentation of a new item, memory would be updated by dropping the old item from the target set and adding the new one. Postle and colleagues further decomposed the potential stages involved in running span into five distinct but coordinated operations: encoding of a new item, discarding of an old item, repositioning of all items within target set, storing and rehearsing the items in the target set (Postle, 2003; Postle, Berger, Goldstein, Curtis, & D'Esposito, 2001).

The processes of encoding, storage and rehearsal of information are commonly involved across tests of ordered memory and have been modelled with a high degree of computational specification as summarised in Section 1.2.2. To recap, serial order can be computationally represented in terms of associations between items and positions (Burgess & Hitch, 1992, 1999, 2006; Oberauer et al., 2012) or using an ordinal system such as a primacy gradient of activation (Farrell & Lewandowsky, 2002; Page & Norris, 1998, 2009). Consequently, items may be output on the basis of position bindings or relative activation levels, respectively. In contrast, in serial-updating tasks, item positions vary throughout a trial, and thus retrieval rarely begins from the first item encoded in the list (Chatham et al., 2011; Juvina & Taatgen, 2007). As such, discarding and repositioning are integral to serial-updating (Postle, 2003; Postle et al., 2001), but the current models of serial recall cannot accomplish these processes.

Three possible solutions to the problem of serial-updating are considered here. These differ primarily in the assumptions made about how serial order is encoded. In one class of models, serial order is represented by pairwise associations between items and position markers (e.g., Burgess & Hitch, 1992, 1999; see also, Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). The simplest assumption might be that, to accomplish serial-updating, these associations change as each new item arrives through a series of unbindings and rebindings to each of the *n* positions. An extension of Oberauer and colleagues' item-removal model might enable such repositioning (Lewis-Peacock et al., 2018; Oberauer, 2018). The no-longer-relevant item could be unbound from the first position to remove it from the target set. The remaining items could be repositioned by successively unbinding them from their current positions and rebinding them to new positions. This way only the relevant number of item-position associations, *n*, would be maintained at any given time-point. This *removal and repositioning* account is illustrated in Figure 1.7a. Rac-Lubashevsky and Kessler (2016) speculated such a model in the case of an *n*-back task, suggesting a gating mechanism to switch between an updating and maintenance mode in working memory (as in Kessler & Oberauer, 2014). This proposed solution would treat serial-updating as a series of individual item-updates.

Models that represent sequences in terms of relative item order rather than absolute positions can potentially address serial-updating directly while avoiding the need for multiple rebindings. Consider the Primacy model, for example, in which serial order is represented as a noisy activation gradient that diminishes across successive list positions, and items are retrieved in order of their activation (Page & Norris, 1998). After *n* items are encoded, each subsequent item could trigger the suppression of the most active (earliest) item. The activation gradient would thus be reset, whereby the second item would have the highest activation becoming the first item to be recalled. In this way, only the latest target set would be rehearsed while older items would be suppressed, enabling the encoded items to be continuously repositioned within the target set. This *suppression* model is illustrated in Figure 1.7b.

Another order-based approach involves associating items with a cyclical representation of a temporal context. Chatham et al. (2011) developed a computational model of *n*-back employing a ring of *n* context-item associations instead of modelling serial order as a linear construct. As each new item arrives, it replaces the oldest item, which is removed and the new item is bound in its position. This automatically implies, in such a cyclical structure, that the position following this latest updated item is now the oldest and would then be replaced in the next updating cycle. Whether cyclical representation is a valid way to conceptualise serial order is unclear. Nevertheless, this *overwriting* model provides a working solution of the serial updating process and is illustrated in Figure 1.7c. Importantly, Chatham et al.'s model incorporates the continuous response procedure characteristic of *n*-back but not end-of-list recall as required in running span. While the same approach of continuous responding could not be applied to running span, Chatham et al. reported that a similar computational architecture could simulate running span performance (p. 3609).

The three proposals considered above are underspecified but provide a first step toward the development of a serial-updating model. The solution based on changing position-item associations calls for one update per position, compared with the order-based solutions in which only the item in the first or last position undergoes a change. In the absence of more detailed specification, the common and only prediction is that there must be a process of updating, possibly specifically beyond the *n*th item in a sequence, to support performance in tasks such as running span and *n*-back. Next, experimental findings with running span (Section 1.3.3) and *n*-back (Section 1.3.4) will be summarised to identify empirical properties that provide important clues as to the operations underlying serial-updating.

1.3.3 Running span

Running span is a task in which participants are asked to attend to a sequence of stimuli and retrieve the last few items from that list. Participants are not informed about the length of the presented sequence and as a consequence, the set of target items relevant for retrieval continuously changes and needs to be updated. It was originally conceived to enable a laboratory study of everyday activities that required continuous monitoring of information from the environment (Pollack et al., 1959, p. 137). Since then, the task has been widely adopted to assess and understand working memory updating (e.g. Botto, Basso, Ferrari, & Palladino, 2014; Bunting et al., 2006; Hockey, 1973; Morris & Jones, 1990; Ruiz & Elosúa, 2013).

1.3.3.1 Running span strategies

Hockey (1973) identified two strategies that participants could use to perform a running span task: passive and active processing (see also, Aaronson, 1968). He showed that patterns of behaviour varied depending strategy participants adopted. Increasing the rate of item presentation improved recall accuracy in participants who were instructed to adopt a passive strategy, but decreased performance in participants using an active strategy. This early finding supported the separation of two modes of input processing during running span. When using the passive mode, participants were said to receive incoming items without engaging in any additional processing or actively attempting to update the recall set. In contrast, an active running span strategy involved attending to and processing incoming items while also keeping track of the target recall set so that only the relevant items are maintained in working memory.

Cowan and colleagues proposed that incoming information during passive listening could be stored as a sensory trace in the first instance and then converted to a categorical form appropriate for recall at the end of the list (Bunting et al., 2006; Cowan et al., 2005; Hockey, 1973). The recall benefit during passive processing at faster rates observed by Hockey indicated that rapid presentation reduced the susceptibility to time-based decay known to characterise sensory memory (see Weems, Winder, Bunting, & Reggia, 2009 for a computational model). For example, if spoken lists are presented, retrieval could be accomplished by relying on representations of the most recent list items in echoic memory, a form of temporary sensory storage to which all spoken inputs have obligatory access (Crowder & Morton, 1969). This proposed reliance on echoic memory remains to be tested; it could be studied by examining whether the use of a post-list suffix (e.g. Morton, Crowder, & Prussin, 1971) impairs the performance in a fast-paced task. A reduction of the recall benefit of the passive strategy observed by Hockey (1973) during rapid presentation would support the hypothesis that passive strategy relies on sensory memory.

The identification of an active strategy that requires ongoing processing and serial-updating of the target set presents a more challenging task. Previously, broad frameworks have attempted to capture it as a drop-and-capture procedure with some suggesting the process of discarding and repositioning items (Pollack et al., 1959; Postle, 2003; Postle et al., 2001). Some computational solutions were considered above (Section 1.3.2). For these, existing computational models of serial order or item updating were extended, and mechanisms that might allow serial-updating in the respective models were speculated. In the following section, experimental investigations of the running span task will be reviewed to suggest possible markers of the serial-updating process and how it may be distinguished from a passive listening strategy.

1.3.3.2 Experimental investigations

Running span performance appears to vary as a function of *n*, i.e. the number of to-be-recalled items (Jahanshahi, Saleem, Ho, Fuller, & Dirnberger, 2008). Using a digit running span, Jahanshashi et al. asked participants to recall the last three, five or seven items. They found that performance was close to ceiling levels of accuracy when participants were asked to recall three items. For recall of five or seven items, accuracy increased up to approximately four items but not further. The researchers suggested that this reflects the maximum capacity in a running span task (p. 24), which was in line with previous estimates of capacity in this specific task (Pollack et al., 1959) and of working memory generally (Cowan et al., 2005).

The length of the presented sequence has a definite but restricted impact on running span performance. Recall accuracy of *n* items is better for sequences containing precisely the number of items to-be-remembered than those containing more than *n* items, with some evidence also suggesting a differential effect across target positions (Salthouse, 2014). In

Salthouse's data, the effect of list length was found for early but not late recall positions. In other words, in lists longer than *n* items, recall accuracy was better for later than early items.

In running span, the sequence length is perfectly correlated with the number of updating steps, meaning that in lists containing more than *n* items, each new item prompts an update. Notably, the effect of list length, and therefore also the effect of the number of updates, does not vary across lists longer than *n* items (Jahanshahi et al., 2008; Morris & Jones, 1990; Postle, 2003; Postle et al., 2001; Salthouse, 2014; see Ruiz & Elosúa, 2013 for similar results using a recognition version of running span). This means that the recall advantage of late compared with early items is independent of whether the list contains $n+1$, $n+2$, or $n+k$ items.

Morris and Jones (1990) explained this invariance in performance as follows. They claimed that performance in running span is only sensitive to the application of the updating process rather than the variable number of updating operations. Updating is not necessary for lists containing only *n* items and therefore, participants show relatively high accuracy across all items. If item presentation continues after *n* items are already encoded, memory updating is required and consequently, performance decreases. These results were interpreted by Morris and Jones in terms of the Baddeley and Hitch working memory model, with the phonological loop maintaining the verbal items and the central executive supporting the updating process (p. 113). The researchers speculated that the executive resources supporting updating are independently deployed when a to-be-updated item is encountered and then replenished within the inter-stimulus interval before the next item onset. It was suggested that the inter-stimulus interval (1 s in their study) reflected the resource recovery rate for updating (Morris & Jones, 1990; see Postle et al., 2001, for a similar argument), but this speculation remains untested.

Recall in running span is characterised by a monotonic serial position function, with accuracy increasing from the first to the last target positions (Bunting et al., 2006; Elosúa & Ruiz, 2008; Hockey, 1973; Morris & Jones, 1990; Palladino & Jarrold, 2008; Ruiz, Elosúa, & Lechuga, 2005; Salthouse, 2014a). The serial position function does, however, vary with the rate of presentation. Bunting et al. (2006) showed that decreasing the pace of presentation from four items to one item per second improved recall at early than late target positions, resulting in a shallower serial position function. This recall advantage at the slow presentation rate was attributed to the employment of an active updating strategy (Bunting et al., 2006; Hockey, 1973). Implicit in this account is the assumption that fast rates deny participants the opportunity to engage in active processing, and the only option available to participants is to rely on passive listening.

An understanding of the temporal boundary conditions for the application of the respective strategies is limited. Studies suggest that a passive strategy is favoured at rates faster than two or more items per second and active updating occurs at rates slower than one item per second (Botto et al., 2014; Broadway & Engle, 2010; Bunting et al., 2006; Collette et al., 2007; Cowan et al., 2005; Elosúa & Ruiz, 2008; Hockey, 1973; Kiss, Pisio, Francois, & Schopflocher, 1998; Morris & Jones, 1990; Postle, 2003; Postle et al., 2001). However, the rate at which participants might shift from an active to a passive approach and the nature of such a shift remain unclear. Results from Hockey (1973, see also Hamilton & Hockey, 1974) show that the performance benefits associated with active updating progressively increase as the rate slows. This result suggests that at intermediate rates, e.g. when an item is presented every 750 ms, participants may choose between the two strategies or even combine them across trials.

There is, however, evidence that cannot readily be accommodated by the account linking running span strategy with the presentation rate. Task performance during a fast and slow-paced running span were highly correlated (Broadway & Engle, 2010). Performance levels in the different rate conditions were also equally correlated with fluid intelligence. Despite observing substantially different levels of performance between rates, the authors interpreted the similarities in correlations as representing the same (passive) cognitive process involved in both a fast and slow-paced running span. Further, Ruiz and colleagues argued that the absence of a primacy effect in running span recall constitutes evidence against the use of an active strategy (Elosúa & Ruiz, 2008; Ruiz & Elosúa, 2013b; Ruiz et al., 2005). They claimed that if updating occurred successfully, positional errors in running span should mirror those in serial recall. In other words, a higher proportion of errors should be found in medial serial positions relative to the early and late target positions, generating standard primacy and recency effects in running span. It was also argued that an active updating account should anticipate a relatively low rate of recall of pre-target items as these items are no longer relevant for retrieval. Contrary to this, evidence suggests a higher recall of items immediately preceding the target set compared with those presented earlier in the sequence (Ruiz et al., 2005). Together, the well-established lack of primacy effect along with the relatively high recall of pre-target items were taken as indications of passive listening in running span (see also, Fiore, Borella, Mammarella, & de Beni, 2012; Palladino & Jarrold, 2008).

This account is inconsistent with previous self-reports describing the slow-paced task as more challenging than its fast-paced equivalent (Bunting et al., 2006). Moreover, analysis of strategy data revealed a variety of active strategies in running span, including rehearsal and grouping (Morrison, Rosenbaum, Fair, & Chein, 2016). Hockey and Hamilton (1977) put forward an alternative explanation for the lack of primacy in running span. They suggested that the standard primacy effect in serial recall originates in a perceptual advantage associated with the first item of an encoded sequence. In running span, however, the first presented item is rarely the first item to be recalled as the sequences are usually quite long. Any items preceding the first target item would eliminate the proposed perceptual advantage and thus remove the primacy component in running span recall. This hypothesis finds empirical support in the re-emergence of a primacy effect in running span when the first target item is also the first item presented in a sequence (Hockey & Hamilton, 1977) or it is signaled by a perceptually distinctive cue, thus reinstating the perceptual advantage (Ruiz & Elosúa, 2013b).

In summary, there are multiple accounts of cognitive processes during running span. Some consider it an effortful updating task, at least under certain temporal conditions such as slow presentation rate. Others have suggested that it involves on a less demanding strategy of passive listening. One way to resolve this debate is to develop behavioural markers that can index active updating and distinguish it from passive listening. In Section 1.4, possible ways to characterise behaviour during serial-updating tasks will be discussed.

1.3.4 *N*-back task

The *n*-back is a continuous-recognition procedure in which a long sequence of stimuli is presented, and participants have to judge whether the current item matches that presented *n* positions back in the sequence (Kane et al., 2007). It was developed independently around the same time as running span to examine the effects of fatigue and ageing on task performance (Kay, 1953, unpublished but cited in Welford, 1958). It is now one of the most popular measures of working memory updating because of its relatively simple stimulus presentation and response requirements. As a result, it is commonly used in studies examining age-related differences in working memory (e.g., Bopp & Verhaeghen, 2018; Hartley, Speer, Jonides, Reuter-Lorenz, & Smith, 2001; Kwong See & Ryan, 1995; Oberauer, 2005; Verhaeghen & Basak, 2005), neuroimaging of working memory abilities and frontal lobe function (e.g., Awh et al., 1996; Cohen et al., 1994; Jonides et al., 1997; Owen, McMillan, Laird, & Bullmore, 2005; Schumacher et al., 1996), and near and far transfer after intensive cognitive training (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Minear et al., 2016; Redick et al., 2013; Soveri, Karlsson, Waris, Gronholm-Nyman, & Laine, 2017). However, despite its apparent ubiquity, relatively few experimental investigations of *n*-back have been conducted to systematically understand the task demands (Kane et al., 2007; Ross, 1966a, 1966b; Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). The following reviews some basic empirical findings from the existing studies investigating *n*-back performance.

Accuracy in performance decreases dramatically and monotonically when the *n* is increased, for example from 1- to 2- to 3-back (Jaeggi et al., 2010; Juvina & Taatgen, 2007; Kane & Conway, 2016; Kane et al., 2007; Mackworth, 1959; Oberauer, 2005; Szmalec et al., 2011). Compared with visual presentation of spatial stimuli, performance appears to be higher for visually presented verbal stimuli (Szmalec et al., 2011) and also higher for spoken presentation of verbal stimuli (Jaeggi et al., 2010). While specific *n*-back conditions, varying in terms of the level of *n* or stimulus modality, were associated with differential performance, they resulted in similar error patterns (Szmalec et al., 2011).

Szmalec et al. proposed that two signals are involved in *n*-back recognition, extending the dual process of recognition initially proposed to account for long-term memory performance (see Yonelinas, 2002 for a review; for similar arguments, see Göthe & Oberauer, 2008; Oberauer, 2005; Oberauer & Vockenberg, 2009; Öztekin & McElree, 2010). One is an explicit *recollection* signal from bound and directly accessible items, containing both item and position information. The other is a *familiarity* signal from all recently encountered items, including those that are no longer bound nor relevant. Both the recognition of targets (items that were recently presented in the *n*-back position) and rejection of non-lure mismatches (items that were not recently presented) were said to be relatively easy since both recollection and familiarity signals provided consistent information. In contrast, lure items created a conflict between the two signals, generating positive familiarity (these items matched a recently presented item) and negative recollection (item from non-target position). Szmalec et al. (2011) proposed that *n*-back performance was particularly prone to such interference and conducted a comprehensive study of interference from different types of lures.

Lures closer to the target position caused greater interference even if they were relatively older. When the relative distance between lures and the target was controlled, lure interference from positions preceding the target was lower compared to that from the positions following the target. Thus, in a *3*-back task, a *2*-back lure (A***B***C*B*) was more interfering than both a *4*-back lure (***B***CDE*B*) and a *1*-back lure (***B****B*). This finding could be attributed to the positional similarity between target and lure positions, in line with the finding that transpositions are more likely to occur between neighbouring positions in serial recall tasks (e.g. Henson, 1998; Hurlstone et al., 2014). However, Szmalec et al. also demonstrated lure interference from semantically related but not identical repeats of the target item. Together, this suggested that lure interference could be due to both a weak, unreliable recollection (positional confusion) and strong familiarity (long-term memory traces with residual activation).

Interestingly, the lure interference effect, as well as the accuracy of target recognition increased when the presentation rate was decreased (Szmalec et al., 2011). In other words, the more time participants were given between successive items, the more accurately they detected matches but the more difficult it was to reject lures. Finally, it was found that the conflict created and potentially resolved during the presentation of lures appears to have downstream (and opposite) effects on subsequent target recognition and lure rejection. While participants tended to miss targets that immediately followed lures (Kane et al., 2007; Moore & Ross, 1963; Ross, 1966b, 1966a), they were more likely to correctly reject lures (Szmalec et al., 2011). Together, this might indicate that the control required to overcome lure interference results in top-down adjustments that decrease the likelihood of responding to lures as well as targets.

As is apparent in the discussion above, the *n*-back literature has typically interpreted results in terms of the embedded processes of working memory such as activated long term memory (related with the familiarity signal) and the focus of attention or region of direct access (related with the recollection signal) (e.g. Oberauer, 2005; Szmalec et al., 2011). In contrast, the studies of running span have often framed results in terms of the multicomponent model of working memory (Baddeley & Hitch, 1974) with claims that performance reflects maintenance in the phonological loop and updating is supported by the central executive (Morris & Jones, 1990).

One study that did extend elements of the multicomponent model to the *n*-back task was carried out by Juvina and Taatgen (2007). Taking an individual differences approach, they outlined two strategies to perform the *n*-back task. One strategy involved high cognitive control such that the latest set of *n* items were rehearsed, in the phonological loop, and older irrelevant items were inhibited, using the central executive. The other strategy described by the authors required low cognitive control and involved neither rehearsal nor inhibition. It was proposed that participants using this low-control strategy automatically 'time-tag' incoming information and rely on an estimate of duration between probe and target item to assist match decisions. The authors showed that relative to the low-control strategy, the high-control strategy was associated with higher *n*-back accuracy and faster reaction times in an independent measure of cognitive control (i.e., the Stroop task). These strategies, and the associated patterns of behaviour, closely resemble the active and passive strategies proposed in the context of running span (Bunting et al., 2006; Hockey, 1973), underlining the argument in this thesis that the two tasks impose similar updating demands on our cognitive systems.

In summary, previous theoretical and empirical studies suggest some degree of executive control is required in both running span and *n*-back. It is assumed that such control is necessary to selectively maintain and update the retrieval set of *n* items in working memory. The experiments presented in this thesis aimed to provide an insight into this serial-updating process. Specifically, the experiments tested if both running span and *n*-back tasks might be characterised by greater demands on executive resources, reflecting the hypothesised cognitive control required to facilitate serial-updating. Another question addressed by the experiments was how the demands differed when high-control, active strategies and low-control, passive strategies were applied to perform the tasks. These questions were explored in four experiments (Chapters 2-5). Section 1.4 briefly outlines the experimental methodologies used in this thesis. It is claimed that these questions can be answered by zooming in on the cognitive processing *during* the stimulus presentation phase of memory tasks, when updating might or might not be occurring, rather than solely relying on the memory performance *after* the presented sequence.

1.4 Studying cognitive processing during item presentation

Most working memory tasks, including running span and *n*-back, can be thought to be composed of at least two phases: the presentation (and encoding) phase and the retrieval phase. Depending on the task, the presentation phase could involve processes such as encoding, maintenance, resequencing, and manipulation of the memory items. Serial-updating is another such process likely to be applied during the presentation phase. However, Kessler and Meiran (2008) maintained that these updating and other encoding processes are usually covert and can only be measured through retrieval accuracy rather than a direct study of the process in isolation. Singh et al. (2018) further added that retrieval measures obscure the individual contribution of updating, maintenance and retrieval processes.

While Singh et al. (2018) and Kessler and Meiran (2008) viewed the limitations described above as intrinsic to the paradigms, the work in this thesis took a different position. It aimed to develop and present behavioural methods that captured the temporal dynamics of updating. Two approaches were adopted. First, demand imposed on cognitive resources during task performance was studied to uncover the time course of updating. Second, the use of specific strategies to perform the task was induced using both experimental manipulation and direct instruction; alongside, strategy reports were also recorded.

1.4.1 Divided attention approach

Attentional (or general executive) involvement during updating can be examined using a divided attention method. This method is based on the premise that central attention or general cognitive resources are limited. If two tasks are performed simultaneously and demand the same resources, then the resources would need to be divided between the tasks and generate a cost in performance in one or both tasks. For instance, Murdock (1966) administered a short-term memory task at the same time as an unrelated but demanding card-sorting task (during the presentation phase). It was observed that performance in both tasks reduced under divided attention conditions relative to single-task performance. The magnitude of impairment in the memory task was associated with the complexity of the card-sorting task, whereby recall was poorer when card-sorting was more demanding. Additionally, the relative performance in one task could be improved by instructing participants to place more priority on it, which showed that the distribution of resources was under volitional control.

Craik and colleagues used this approach to test if the same mechanisms underpinned both encoding and retrieval in long-term memory (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). They studied performance when attention was divided during encoding or retrieval by the concurrent application of a choice reaction time (CRT) task. Any cost in dual-task performance relative to single-task conditions was associated with the distribution of resources between the memory task and the CRT task. In this way, dual-task costs were used to index the resource demands of encoding and retrieval. The authors reasoned that if the encoding and retrieval processes are similar and rely on the same resources, then the performance costs should be similar regardless of the phase in which attention is divided. Conversely, a differential impact of divided attention would constitute evidence that the two processes were different. Results showed that there were significant costs in both memory and CRT performance when attention was divided at encoding. However, divided attention at retrieval had relatively little impact on memory performance and markedly greater impact on concurrent CRTs. Also, memory performance was sensitive to changes in relative priority between tasks at encoding but not at retrieval. The relatively slower CRTs during retrieval suggested that it demanded more resources than encoding and was resilient to the effects of divided attention. Also, the allocation of resources away from the CRT task during retrieval appeared obligatory and outside strategic control. This study thus effectively used the divided attention approach to infer that the processes of encoding and retrieval are distinct.

This approach was advanced in later studies by Naveh-Benjamin and colleagues to index resource demands in a precise, event-related manner (Naveh-Benjamin, Craik, Guez, & Dori, 1998; Naveh-Benjamin, Craik, Perretta, & Tonev, 2000). In one study, a tracking procedure was used in which participants were required to track a target on the screen, and the spatial distance between their position and the target was continuously recorded (Naveh-Benjamin et al., 1998). This continuous spatial distance function was then segmented into *events* (the interval around the encoding or retrieval of an item) and *between-events* (the remaining intervals during the course of the trial). These measures provided temporally precise indices of the resource demands of encoding and retrieval, replicating the general observation by Craik et al. (1996) that the two processes are different. Such measurement added temporal specification as encoding and retrieval could be distinguished not only at a trial level but also on the basis of demands imposed during specific events and in the periods between these events.

This approach was recently extended to the study of maintenance processes within working memory (Thalmann, Souza, & Oberauer, 2019). In this study, a verbal short-term memory task was administered with a 10-second delay between encoding and retrieval. During this delay, participants were asked to maintain the encoded words using articulatory rehearsal or semantic elaboration or under articulatory suppression. Additionally, a visuospatial CRT task was administered during the delay period, such that attention was divided between the CRT task and the maintenance processes in working memory. The CRTs in the concurrent task were used as an index of the resource demands of the three maintenance conditions. As in the study by Naveh-Benjamin et al. described above, the authors examined resource demands over time, although with less temporal specificity. The 10-second delay period was divided into five segments and average CRTs were computed for each segment. In this way, the authors determined that the articulatory rehearsal demands cognitive resources over the entire course of the delay period. The evidence was less consistent for the demands associated with semantic elaboration and maintenance under articulatory suppression imposed minimal resource demands.

The divided attention approach as such provided a valuable methodological avenue to examine the resource demands of serial-updating. It was therefore employed in three experiments presented in this thesis to examine the demands during running span (see Chapters 2-3) and in *n*-back (see Chapter 5). In each experiment, a visual CRT task was applied concurrently with the respective updating task, and the CRTs were used to index the demands imposed by the memory task. As in the studies above, the experiments presented in the following chapters capitalised on the continuous nature of the concurrent CRT task, thus generating a continuous metric of the resource demands of updating. In this way, the cognitive demands of updating were tracked at a fine temporal resolution. The experiments in this thesis represent the first examination to date of how updating and non-updating tasks vary in their cognitive demands.

1.4.2 Strategy reports and induction

Another approach to gaining insight into a task is to focus on the strategies employed by participants to perform the tasks. Self-report data has previously demonstrated extensive heterogeneity in strategy use both within individuals and across working memory tasks, including running span (Morrison et al., 2016). It has also been successful in accounting for individual differences as well as training-related gains in performance. For instance, a study found that high-performing individuals used strategies such as visual imagery or grouping rather than simple item repetition and also displayed greater flexibility in their strategy use (Dunlosky & Kane, 2007). These data are particularly informative in complex working memory tasks that could be approached in multiple ways (Norris et al., 2019). Given its potential value, self-reported strategy data was also collected in work presented in this thesis. It was sampled through both structured (Chapter 3-5) and open questions (Chapter 4).

A second approach to studying the cognitive operations used to perform complex tasks is by inducing participants to use particular strategies. Implicit strategy induction can be carried out by employing conditions that favour some strategies over others. For example, Bor et al. reported that participants could be steered toward chunking or grouping strategies if the input stream contained mathematically structured sequences (Bor, Cumming, Scott, & Owen, 2004). In the context of serial-updating, Bunting et al. (2006) found that different presentation rates in running span resulted in differential reliance on passive and active strategies. The same approach was adopted and extended in this thesis (Chapter 3).

Strategies can also be manipulated through explicit instructions (Carretti, Borella, & De Beni, 2007; McNamara & Scott, 2001; St Clair-Thompson, Stevens, Hunt, & Bolder, 2010; Turley-Ames, 2003). Norris et al. (2019) used this approach to provide data that challenged the enduring assumption that backward recall is performed using a peel-off strategy. The authors showed that the response latencies were markedly different between backward recall trials without any strategy instructions compared to trials in which participants were specifically instructed to use a peel-off strategy. This finding indicated that participants did not spontaneously employ a peel-off strategy to complete backward recall in the absence of instructions, contrary to previous assumptions (Anders & Lillyquist, 1971). A similar technique was applied by Hockey (1973) to study running span, permitting a direct comparison of behaviour. In his study, participants were instructed to adopt either a passive or an active strategy while other task conditions, such as the rate of presentation, were held constant. This thesis adopted Hockey's method of strategy induction to study the impact of employing different strategies on processing times in a self-paced running span (Chapter 4).

To conclude, this thesis addressed the critique levied by Kessler and Meiran (2008) at previous investigations of serial updating. The objective was to combine the study of resource demands during the encoding (and updating) phase of memory tasks and apply experimental control over strategy use, in order to study cognitive processing during serial updating.

1.5 Main aims and thesis structure

Studies have demonstrated that working memory updating plays an essential role in everyday cognition and, unlike other executive functions, predicts general intelligence. Recent investigations have illuminated cognitive operations underlying item-updating (Section 1.3.1), but relatively little is known about how updating is achieved on the basis of serial order. A few potential computational solutions that could simulate serial-updating are discussed in Section 1.3.2, but pending implementation these do not clarify the mechanisms involved. A review of studies involving serial-updating tasks in Sections 1.3.3 and 1.3.4 raises several possible, untested characteristics of the process. Therefore, the overarching aim of this thesis is to provide greater specification of serial-updating than currently available, with a particular focus on examining the resource demands over the course of the updating activity.

The first study of this thesis is presented in Chapter 2. This experiment aimed to develop an index of the resource demands associated with serial-updating. For this, a divided attention approach was used, in which an updating task (running span) and a choice reaction time task were administered simultaneously. The costs in the choice reaction time task provided a continuous measure of cognitive demands throughout the running span task. These demands were also contrasted with two comparable short-term memory tasks that did not involve updating. This comparison enabled the identification of the features in the demand profiles specific to running span, thereby suggesting behavioural markers of the process of serial-updating.

In the second experiment, detailed in Chapter 3, the aim was to examine if the resource demands identified in the first experiment are sensitive to temporal conditions of the task. If the demands reflect an updating process, they should be replicated during slow presentation conditions that favour active updating strategies. In contrast, fast-paced tasks, associated with the use of passive strategies and known to prohibit updating, should not exhibit the same resource demands found in the first experiment. To test this hypothesis, the divided attention approach was employed. The choice reaction time task was applied at the same time as the running span task, and three conditions were used in which the rate of presentation in the running span task was varied. In this way, the resource demands in running span at three presentation rates were mapped, clarifying the temporal boundary conditions of the updating process.

The third experiment is elaborated in Chapter 4. This study built on the findings of the first two experiments that suggested that active updating in running span is time-consuming and demanding while a passive approach is relatively effortless. The primary aim of this experiment was to distinguish the two strategies in terms of processing times. Participants were instructed to use either an active or passive approach in the running span task. A self-paced procedure was used in which participants regulated the presentation rate of the memory items rather than using a rate determined by the experimenter. These item presentation times chosen by the participants were measured over the course of the trial and compared in the two strategy conditions. A secondary aim of the study was to understand how participants solve the task using either active or passive strategies. For this, extensive self-report strategy data were obtained and analysed. Taking a convergent operations approach, this study examined the hypothesis that active updating requires time and cognitive resources and is distinct from passive listening.

The aim in the fourth experiment, presented in Chapter 5, was to investigate the extent to which the profile of resource demands in running span is generalisable to a similar but not identical task. In this study, the focus was on the *n*-back task that involves recognition decisions in addition to serial-updating. The aim was again to chart resource demands using the divided attention approach through the course of the *n*-back task and to then compare the resource demands in *n*-back with those in running span. A similarity in features in the demand profiles of both tasks could indicate the shared serial update process while the dissimilarity could indicate processes unique to each respective task.

Finally, in Chapter 6, a general discussion is presented. It summarises the outcomes of each experiment and elaborates the theoretical implications of the work presented in the thesis. It also identifies the study limitations and avenues for potential research in the future.

Chapter 2. Cognitive demand during updating in running span

2.1 Overview

A wealth of empirical and computational research has advanced our understanding of how serial order is preserved, when information is encoded and retrieved in working memory. It is, however, unclear how a representation of serial order could provide the basis for updating contents in working memory. Running span is a valuable experimental tool to investigate such serial updating of working memory. In this task, participants are presented with long memory sequences containing a variable number of items and are asked to recall the last *n* items in serial order once the sequence ends. Participants must therefore not only encode and rehearse the items but also update the target set so that only *n* items are maintained in memory at any given point in time. It thus requires the encoding and retrieval of information in serial order, as in standard serial recall tasks, with the additional demand of memory updating. Comparing serial recall and running span therefore provides an effective means of isolating the distinctive updating component of running span whilst controlling the serial recall demands common to both paradigms.

Experiment 1 employed this approach to characterise the updating process. A divided attention method was used to assess the demands on cognitive resources as participants completed running span. For this, a choice reaction time (CRT) task was performed at the same time as running span, with the CRTs in the concurrent task used as an index of the resource demands of running span. The resource demands of two additional serial recall tasks that did not require serial updating of working memory were also measured in the same way. The demand profiles were then compared between the three tasks, allowing the identification of the demand characteristics specific to the running span task.

The results of Experiment 1 and 2 reported in this and the following chapter have been accepted pending minor corrections as an article to the Journal of Experiment Psychology: Learning, Memory, and Cognition.

2.2 Introduction

Working memory (WM) allows individuals to temporarily hold, manipulate and process information to support goal-directed behaviour (Baddeley & Hitch, 1974). A common assumption of WM models is that it is a capacity-limited system that can hold only a restricted amount of information at any given time (Baddeley, 2012; Baddeley & Hitch, 1974; Cowan, 1999, 2005). Because of this limitation, a process of WM updating is hypothesised that acts as a filter (Ecker, Oberauer, et al., 2014; Kessler & Meiran, 2008). Updating preserves the information that is task-relevant while letting go of the information that is distracting or outdated. For instance, consider a customer placing an order for one medium and one large red wines in a bar. If the customer then changes their order of the large red to a large white wine, the bartender's internal representation of the order would be updated and reflect the same change. In this way, updating ensures that the contents of WM are aligned with ongoing tasks and relevant information is accessed efficiently. The process of WM updating has previously been associated with fluid intelligence (Chen & Li, 2007; Friedman et al., 2006) and is important in everyday cognition, such as arithmetic (Deschuyteneer et al., 2006), navigation (Garavan, 1998; Gugerty, 1997), reasoning (Carpenter et al., 1990), reading (Carretti et al., 2005; Palladino et al., 2001) and problem-solving (Passolunghi & Pazzaglia, 2005).

Previous work demonstrated that WM updating involves a two-component process in which irrelevant information is first actively removed from memory and then replaced by newly-relevant material (Ecker, Oberauer, et al., 2014; Oberauer, 2018). The active removal mechanism occurs on an item basis, involving sequential scanning followed by un-learning to remove the outdated item when it is encountered so that it can be substituted by the new item (Section 1.3.1.1 contains a more detailed discussion). This proposed mechanism successfully accounts for behaviour when *individual* items are updated while the rest remain unchanged (Ecker, Lewandowsky, et al., 2014; Ecker et al., 2010; Ecker, Oberauer, et al., 2014; Kessler & Meiran, 2008). It is unclear if the same mechanism can account for serial updating activities in which the sequential positions of *all* items change rather than the items themselves.

Running span is a task that requires serial updating. In this task, participants are presented with long memory sequences containing a variable number of items and are asked to recall the last *n* items in serial order once the sequence ends. Participants must therefore not only encode and rehearse the items but also update the target set so that only *n* items are maintained in memory at any given point in time. The procedure to present stimuli and test memory in a running span task resembles the well-studied paradigm of immediate serial recall. However, the running span task imposes an additional demand to continuously change the set of target items and thus provides a useful experimental tool to investigate WM updating.

Previous studies have identified a number of key characteristics of running span performance that have guided our understanding of how serial updating may be accomplished (Section 1.3.3.2). First, active updating during running span is a slow process. When participants were instructed to process items during the task actively, task performance improves with slower rates of presentation (Hockey, 1973). The benefit associated with the slow presentation rate is found even when no explicit strategy instructions were provided (Bunting et al., 2006). This implies that active updating is a time-consuming process and cannot be effectively applied when information is presented rapidly. Second, updating in the running span is a highly demanding process. Participants found a slow-paced running span requiring active WM updating more challenging than its fast-paced equivalent (Bunting et al., 2006). Consistent with this, performance in a slow-paced running span was impaired when resources were engaged by a concurrent executive task (Botto et al., 2014).

Morris and Jones (1990) found that recall accuracy in running span was higher for short lists (= *n* items long) requiring no updating, compared with lists that contained more than *n* items and thus required updating. This suggests that updating demands are restricted to lists longer than the target recall set of *n* items. The same study also found that the number of updating steps in running span did not impact recall performance (see also, Jahanshahi, Saleem, Ho, Fuller, & Dirnberger, 2008; Postle, Berger, Goldstein, Curtis, & D'Esposito, 2001; Ruiz & Elosúa, 2013; Salthouse, 2014), indicating that the resources that support memory updating in running span might be deployed and then replenished within the interval between successive stimuli. The inter-item interval may therefore reflect the recovery rate for updating resources (Morris & Jones, 1990; Postle et al., 2001). The minimum amount of time required for such recovery is unclear as the interval tends to vary across experiments but appears to be at least 1 s duration as used in the study by Morris and Jones (1990).

Although running span is widely used as a WM updating task (e.g. Botto et al., 2014; Bunting et al., 2006; Hockey, 1973; Morris & Jones, 1990; Bradley R Postle, 2003), its utility has been questioned (Elosúa & Ruiz, 2008; Ruiz & Elosúa, 2013a; Ruiz et al., 2005). Ruiz and colleagues have argued that the consistent lack of a primacy effect in running span suggests that participants are not actively updating WM, relying instead on passive listening to support recall (for a longer consideration of this argument, see Section 1.3.3.2). However, the absence of a primacy effect does not necessarily imply that running span does not involve WM updating. Hockey and Hamilton (1977) suggested that the primacy effect may indicate the perceptual advantage granted to the first item presented in a sequence. The primacy component may thus disappear if the first item to-be-recalled differs from the first presented item, which is often the case in long running span lists. Thus, recall performance may not be an appropriate indicator of the presence, or nature, of the updating process. This highlights a need for an alternative method to examine the cognitive operations of updating *during* rather than *after* the presentation phase.

Dividing attention by introducing a concurrent task during encoding is one way of assessing the demands of running span on general attentional resources. That logic is that two cognitive tasks performed simultaneously should compete for the same resources. This should thus decrease the resources available for each task, impairing performance in one or both tasks (e.g. Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Murdock, 1966; Thalmann, Souza, & Oberauer, 2019). For example, Murdock (1966) measured the short-term memory capacity of participants while they were simultaneously sorting cards. Both short-term memory and card-sorting performance reduced in the dual-load condition compared with the respective single-load conditions. It was also found that memory performance showed greater impairment when combined with card-sorting based on four suits instead of two colours. Increasing the complexity of card-sorting thus appeared to increase the demands imposed by the card-sorting task while reducing the resources available for, and performance in, the memory task. These findings suggest that performance in one task is sensitive to, and provides a useful index of, the demands of the other task when two tasks are performed concurrently.

Craik and colleagues used the divided attention approach to determine if encoding and retrieval processes in long-term memory imposed the same demands (Craik et al., 1996). In their study, participants encoded and retrieved words (or word pairs) in the memory task while also completing a choice reaction time (CRT) task. The CRT task was applied either during the encoding phase or the retrieval phase of the memory task. It was argued that the costs in task performance due to the simultaneous application of the memory and the CRT task would provide an index of the demands during encoding and retrieval processes. The data indicated a marked difference in the pattern of dual-task costs during the two processes, suggesting that encoding and retrieval are distinct in their resource demands and hence their underpinning mechanisms.

The utility of this approach was extended by Naveh-Benjamin and colleagues by combining the memory tasks with a continuous spatial tracking task (Naveh-Benjamin et al., 1998, 2000). In this study, participants had to continuously track the spatial location of a target while also performing the memory task, i.e. the task of main experimental interest. Performance in this concurrent task could be queried at any point, thus yielding a continuous measure of resource demands that spanned the entire course of the memory task. This imparted temporal specificity that allowed an event-related analysis of demands of encoding and retrieval (for a more elaborate discussion of the results, see Section 1.4.3).

The divided attention approach has been used more recently to understand the involvement of executive resources in the maintenance of verbal content in WM (Thalmann et al., 2019). In this study, the researchers administered a verbal WM task with a delay period between encoding and retrieval and instructed participants to use different strategies to maintain the items in WM. A CRT task was performed during the delay phase, which was segmented into five parts. The CRTs indicated the cognitive demands imposed by the different maintenance strategies within each two-second segment and these were compared across the strategy conditions. The results indicated that articulatory rehearsal but not semantic elaboration is supported by general cognitive resources over the course of the delay period.

Experiment 1

The mechanisms and time course of serial updating in running span are not yet well understood. The work presented in this thesis took a novel approach of dividing attention as participants performed running span in order to cast light on the task. Experiment 1 is the first study to examine resource demands during the process of serial updating in this way. For this, running span was combined with a continuous CRT task involving a four-alternative forced choice with RTs from this task used as a metric of the time course of resource demands during running span trials.

Although the study was primarily exploratory in nature, a number of possible outcomes could be identified. These are represented schematically in Figure 2.1. First, it may be that an unvarying level of demand persists across the running span trial. This outcome would be expected if participants adopt a running span mode in which the target recall set is continuously monitored with each incoming item in the memory sequence. Another possibility was that an updating mode is initiated only when the first to-be-updated item is encountered and additional resource demands are imposed to support updating from this point in the trial. This hypothesis is in line with previous speculations by Morris and Jones (1990) who suggested that updating is activated only when an encoded list is longer than the target number of items to be recalled (see also, Postle et al., 2001). If this were the case, the demands would be expected to specifically increase following the $n+1$th position in the encoded sequence, where n is the number of to-be-recalled items. A third possibility was that each updating cycle triggers a short-lived serial update process evident within the item interval but not beyond. In this case, resource demands would rise with each update item in the sequence. After a certain duration, these demands may fall to a baseline level which would indicate the recovery of resources as speculated by Morris and Jones (1990; see also, Postle et al., 2001). As the hypotheses outlined above do not preclude each other, a final consideration was that a combination of these alternatives may be observed. In this case, heightened resource demands could be found across all update items in addition to a localised peaks in the intervals between successive update items. The divided attention approach, therefore, provided a powerful methodology to consider these alternative hypotheses and examine the precise costs of time-locked events within running span.

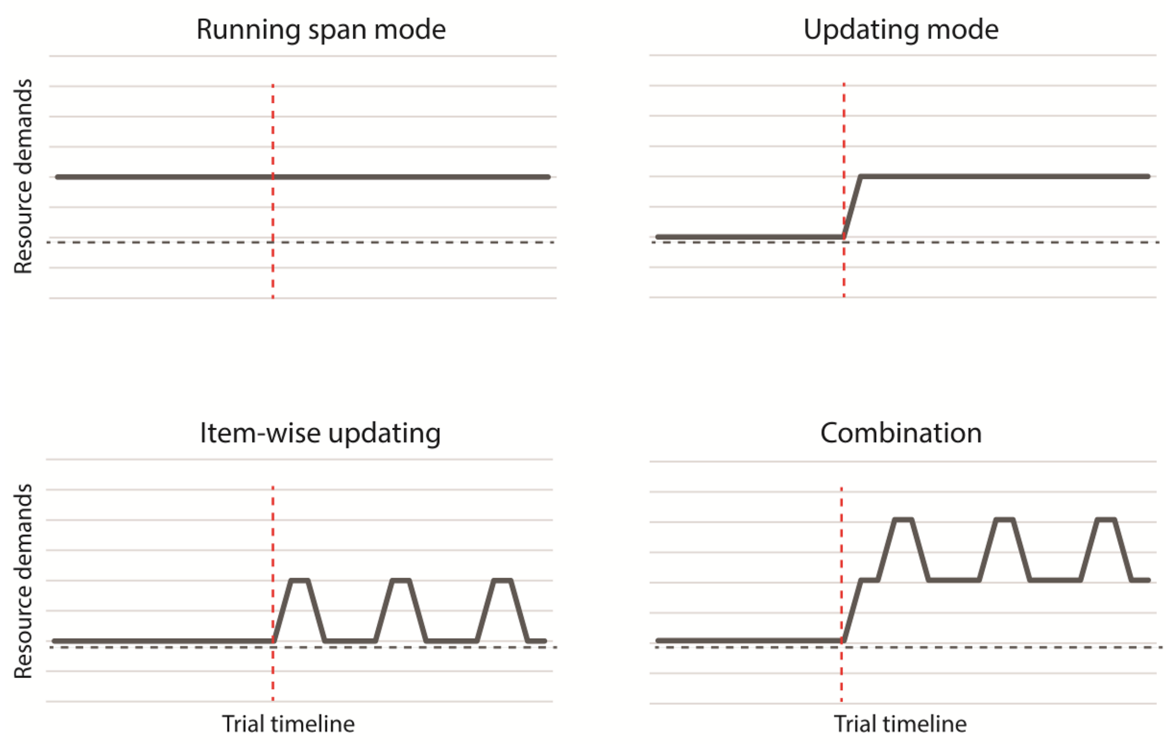

Figure 2.1 Four alternative hypotheses considered in Experiment 1. The vertical dashed red line denotes the onset of the first updating item at position *n+1*, where *n* is the number of items to be recalled in running span. The horizontal dashed black line denotes the baseline resource demands for standard serial recall without updating. See text for explanation.

Clearly, the demands imposed by running span reflect multiple component processes that include the encoding, maintenance, updating and retrieval of item and order information. It was therefore essential to compare the demands in running span with those in other tasks that entailed similar processes without serial updating of WM. The experiment also measured the resource demands of two serial recall tasks, making it possible to isolate the demands specific to the serial updating process.

Simple span, the first comparison task, required standard serial recall with no updating of items or their sequence. It was hypothesised that higher resource demands would be observed in running span due to its additional serial update requirements. Similar demands in both tasks would instead favour the use of passive listening strategies to perform running span, in line with suggestions by Ruiz et al. (2005).

The second comparison task was a modified version of simple span that required a change in WM representations prior to recall but no item-wise serial updating. Whole lists were presented and followed by cues either for recall or to start encoding afresh. The cues therefore induced a complete reset of the WM contents. This memory reset differed from the serial update in running span in several ways. The memory reset required a complete change of stored material, whereas the serial update needed specific changes to the target set (i.e. eliminate the oldest item from the target set, add the newest item, and reposition the other items within the target set). Additionally, memory was reset periodically between component target sets in modified span trials while memory was updated continuously after each update item in running span. Another distinguishing feature was that the reset event was driven by external cues, while the update event was not cued and thus internally driven. Previous studies of item updating have found that changing an entire set of encoded items is markedly faster than changing a subset (e.g., Ecker, Oberauer, et al., 2014; Kessler & Meiran, 2008). Ecker and colleagues suggested that representations in WM may be eliminated using a single, rapid process, whereas partial updating of WM proceeded separately for each to-be-updated item (Ecker, Oberauer, et al., 2014). Based on the task distinctions as well as Ecker et al.'s findings, it was anticipated that the reset events in modified span would impose fewer demands than the update events in running span. The difference in demands associated with simple span and modified span were also explored to identify the costs of shifting from maintaining a memory set to restarting the encoding of an entirely new one.

In summary, Experiment 1 aimed to examine the time course of demands during serial updating in the running span task and compare it to the demands associated with two comparison serial recall tasks that required either no change or a complete reset of WM. This study presented an important first step towards developing a process model that provides greater specification of serial updating than current theories. Dividing attention by performing a concurrent RT task enabled the quantification of the resource demands of running span across the full time course of the trial.

Pilot studies

When two tasks are undertaken simultaneously, the impact on the performance in a task is known to be sensitive to the complexity of the concurrent task (e.g., Murdock, 1966). It was therefore necessary to ensure that the difficulty levels across the three memory tasks were comparable, as it would rule out task difficulty as a source of variation in the resource demands between tasks.

Two pilot investigations were conducted to achieve comparable difficulty levels (Appendix B). Task difficulty was adjusted by varying the number of target items to be recalled (*n*). Recall accuracy in the single load condition was measured for each memory task and the *n* associated with 70-85% accuracy was used in the main experiment. In running span, *n* was 4 (associated with 83% accuracy); in both modified span and simple span, *n* was 7 (associated with 73% and 82% accuracy respectively). Given the parity in recall levels, task-related differences observed in the main experiment were unlikely due to overall task difficulty and thus reflected specific differences in cognitive demands across tasks. The pilot studies were also used to test the divided attention approach (details about task description and procedure are presented in the following section). Compared with single load conditions, recall accuracy was impaired in the memory task and RTs increased in the concurrent CRT task in each dual load condition. These dual-task costs were consistent with previous studies using a similar combination of a visual-motor CRT task and a verbal-spoken memory task (e.g. Craik et al., 1996; Thalmann et al., 2019).

2.3 Method

2.3.1 Participants

Ninety-two native English speakers aged 18-40 years with normal or corrected-to-normal vision were recruited to participate in the experiment. Complete data were recorded for 90 participants (68 female, 22 male, mean age= 24.38 years, *SD*= 4.04 years). Participants were recruited via the volunteer panel at the MRC Cognition and Brain Sciences Unit, University of Cambridge and through electronic advertisements in Cambridge. They had normal or corrected-to-normal vision and hearing and did not report any psychological, psychiatric or neurological disorders that would impede their participation in the experiment. Informed consent was obtained in accordance with ethical approval from the Cambridge Psychology Research Ethics Committee (PRE 2016.066; Appendix A), and time and travel costs incurred by the participants were compensated.

2.3.2 Procedure

The study used a 3x2 mixed factorial design. Task was a between-group factor with three levels consisting of the three WM tasks: simple span, modified span, and running span. Participants were randomly assigned to one of three groups completing different memory tasks respectively. Attentional load was a within-subject factor with two levels: single and dual load. Also, all participants completed a digit span task for an assessment of verbal short-term memory capacity.

At the beginning of the session, informed consent from the participant was acquired. The digit span task was then administered, followed by the three experimental tasks: the choice reaction time (CRT) task, the respective working memory (WM) task, and the dual load condition in which both CRT and WM tasks were applied concurrently. These tasks were performed using a blocked design with a fixed order (CRT, WM, and dual load) implemented for one block of practice trials and five blocks of experimental trials. The experimenter was present during the practice trials to ensure that the participants understood and followed task instructions. All tasks were performed in a sound-proof room and completed within one session, typically lasting between 1.5 to 2 hours. All tasks described in this and the following chapters were designed and presented on a PC using MATLAB 2014a (The Mathworks, Inc.) and Psychtoolbox-3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

2.3.3 Tasks

2.3.3.1 Memory tasks

The three memory tasks (Figure 2.2a) were completed by separate groups of participants respectively. The specific instructions, list lengths, and size of target recall set varied across the tasks as detailed below. The number of target items to be recalled in each task was determined based on pilot data as summarised earlier and reported in Appendix B. The WM load associated with an accuracy between 70-85% in each pilot memory task respectively (single load condition) was selected.

Running span. Participants were required to recall the last four items of the presented list in correct serial order in each trial. Trials started with a one-second alerting tone and contained a sequence of 4 to 12 items. The length of the list in an upcoming trial was unknown to the participants. An experimental block contained ten trials of varying lengths presented in a pseudo-random order, such that each list length was presented once and lists containing eight items were presented twice per block. Participants thus completed fifty trials over five blocks.

Simple span. Each trial started with a one-second alerting tone and contained seven items. Participants attempted to recall the entire sequence in order. Fifty lists were presented over five experimental blocks.

Modified span. Participants were presented with sequences of letters periodically interspersed with tone cues that indicated the start of the set of items they were to remember. The sequences contained 7, 14, 21, or 28 items. Therefore, each memory sequence comprised between one and four target sets of seven items. Items in positions 8, 15, and 22 were considered update items as they were presented in the context of lists longer than seven items. Even though they represented the start of a target set, this was considered to be different from the item at position 1. While the first item was always the start of a new sequence, update items were the start of later target sets, preceded by one or more target sequences. A one-second tone was presented before each update item to alert the participants to start encoding a new set of items. Participants were asked to recall the last target set immediately following the end of the sequence. The number of items to be recalled was always fixed at seven, as in simple span, but the length of each list was unknown to the participants. Each list length was presented twice in a pseudo-random order in a block. There were forty trials presented over five blocks.

List generation and presentation. Identical protocols were employed to generate and present memory lists in all three memory tasks. Lists were generated on the basis of the following rules: (a) 20 consonants from the English alphabet were used as stimuli ('W' was excluded), (b) letter repeats were only allowed after every 7 items, (c) consecutive presentation of phonologically similar letters was not permitted, (d) three or more letters in alphabetic order could not be presented. A typical sequence containing nine items was *D, S, P, Y, R, L, G, K, D*.

The letters were spoken by a male speaker of British English and were recorded at a sample rate of 44.1 kHz. The sound files for each letter centred so that they were 800 ms long and the location of the letter was adjusted so that the letters in the presented sequence sounded evenly spaced (Morton, Marcus, & Frankish, 1976). The sound files containing letters were concatenated with silent intervals of 1600 ms. These 2400 ms files were sequentially assembled to produce auditory stimuli of varying length according to the rules listed above. The auditory stimuli containing memory items were presented to participants using headphones (Sennheiser HD 280 PRO II) at a rate of 2400 ms per item (i.e., 800ms for item duration followed by a silent interval of 1600ms). Such a slow presentation rate was chosen to encourage the use of active updating strategies in running span (Bunting et al., 2006; Hockey, 1973). Each sequence was preceded by a one-second tone cue to alert participants to the start of the trial. A white fixation cross was displayed on a grey screen throughout the trial. The duration of the presentation phase was determined by the length of the memory list and was immediately followed by a response phase for a maximum duration of 20 s.

Recall recording and scoring. Participant response in the form of spoken serial recall was digitally recorded using a table microphone. These recordings were then manually transcribed by the experimenter. Recall accuracy was measured in terms of the proportion of items recalled in the correct serial position. Participants were allowed to skip recall of a letter that they could not remember by saying "blank" or "space" at the respective serial position. For example, if the target set comprised of *S, L, J, G*, a response "S blank J G" would receive 75% credit, while "S J G" would receive only 25% credit.

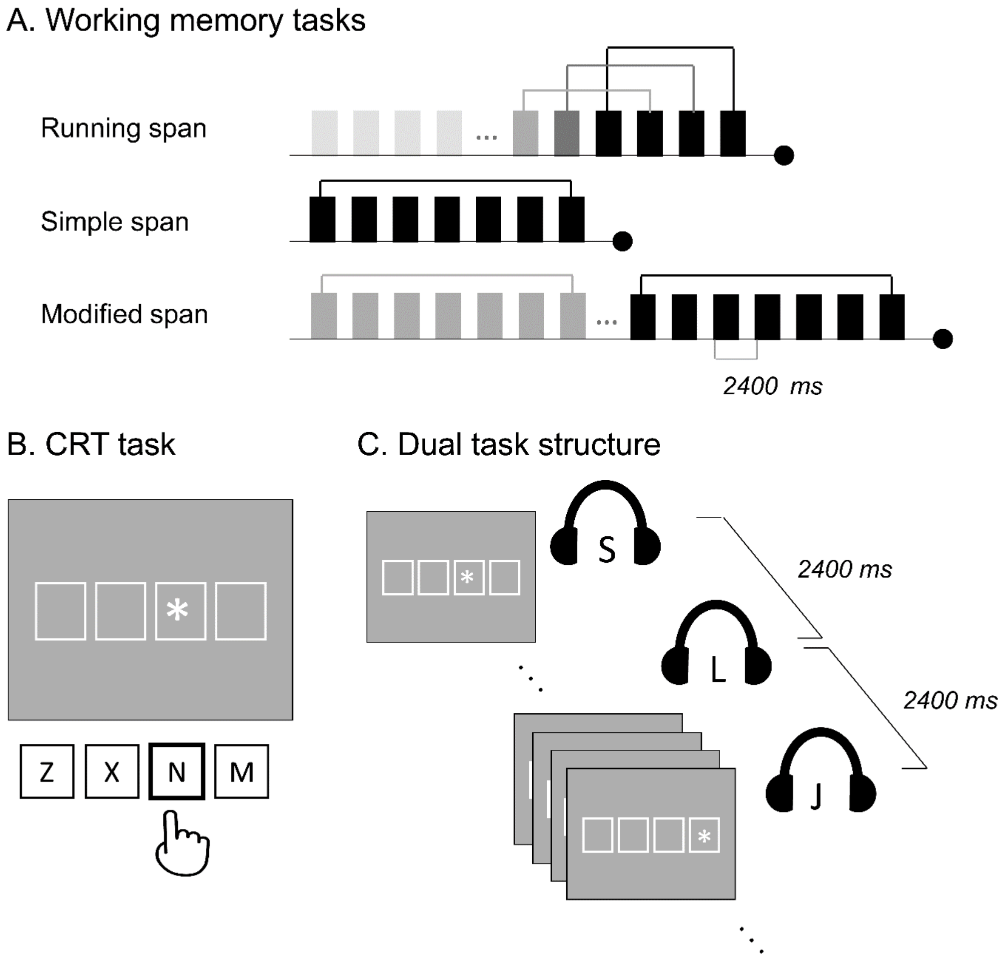

Figure 2.2 Task structure in Experiment 1. (A.) Schematic of a memory list for each working memory task. Memory items (marked in rectangular boxes) were presented sequentially at a rate of 2400ms per item (800ms for the item, followed by a 1600ms silent interval) using spoken presentation. List length varied in modified span and running span. The later items in the list were relevant for recall (marked in black, varied as per task), while earlier items were not (grey). The brackets above items denote the items in the same target recall set at a given time-point. (B.) The continuous choice reaction time (CRT) task. Four square frames were presented on the screen corresponding to four response keys. Participants pressed the key corresponding to the frame containing the star. A CRT stimulus was presented immediately following the response to the previous stimulus. (C.) Dual-task structure with a simultaneous application of (auditory) working memory and (visual) CRT task. Memory items presented every 2400ms; CRT task was participant paced, so the number of CRT stimuli presented between each memory item varied across participants, contingent on their RTs.

2.3.3.2 Choice reaction time task

In the choice reaction time (CRT) task, an asterisk was presented in one of four possible locations on the screen (represented by square frames arranged horizontally in the centre of the screen; Figure 2.2b). The location of the asterisk on any given trial was pseudo-randomly determined such that the asterisk did not appear in the same location over two consecutive trials. Participants were asked to press the key corresponding to the location of the asterisk as quickly and accurately as possible. Responses in this task involved key-presses of one of four keys *Z*, *X*, *N*, and *M*, corresponding to the four spatial locations on the screen. These were mapped on to the first two fingers of the right and left hands. The CRT task was self-paced and continuous. So, the onset of the next stimulus array immediately followed each participant response. Both reaction time (RT) and accuracy were recorded in this task.

2.3.3.3 Dual load condition

In the dual load condition (Figure 2.2c), the WM and CRT tasks were simultaneously administered. The protocol for presentation and generation of stimuli in this task was identical to the respective single tasks. In each trial, the visual onset of CRT stimulus was synchronised with the auditory onset of the respective WM task. For the remainder of the trial, CRT stimuli were presented successively during the ongoing presentation of the memory sequence and terminated as soon as the recall phase of the WM task commenced. Participants attended the memory items for subsequent recall in the respective WM task while also responding as quickly and accurately as possible in the CRT task using keypresses. It was emphasised that both tasks should be treated with equal priority.

2.3.3.4 Digit span task

Verbal short-term memory capacity was measured using a digit span task. In this task, digits (0 to 9) were presented sequentially in a pseudorandom order such that a digit could be repeated within the same sequence only after nine items. Each digit was presented for a duration of 1000 ms in the centre of the screen and followed by a 1000 ms inter-stimulus interval (ISI; blank screen). At the end of the list, the participants were prompted to recall the digits in serial order. Responses were indicated using mouse clicks on a response keypad displayed on the screen. The task commenced with lists containing four items. At each span level, four trials were presented. If participants responded with 75% accuracy at any given level, they advanced to the next level (i.e. the list lengthened by one item). The task was discontinued once performance failed to meet this condition. Span was recorded as the longest list length at which participants correctly recalled three or more trials.

2.3.4 Statistical power

As there was little previous research using the divided attention approach as a method of temporal analysis of serial recall, *a-priori* power analyses could not be performed. Previous studies investigating running span were used to inform the present sample size of thirty per group (Bunting et al., 2006; Ruiz & Elosúa, 2013a). Thalmann et al. (2019) used a similar sample size in their study using the divided attention approach to understand maintenance processes in short-term memory, albeit without the detailed temporal analysis undertaken in the present study. The number and length of trials yielded over 100 data points in each bin per individual in the analysis at the item level. The post hoc power was greater than .90 for all significant effects reported below, and in most cases it was greater than .95.

2.3.5 Analysis plan

Reaction time. Responses initiated faster than 200 ms or due to accidental holding down of a response key from the previous trial were excluded in both single and dual load conditions of the CRT task (Van Zandt & Townsend, 2014). Programming constraints caused the CRT task to abruptly stop upon reaching the end of the memory list in the dual load condition, which truncated the recording of any CRT responses that may have followed. Therefore, dual CRT responses associated with the final item in each memory list across all tasks were removed. The data were then trimmed by first removing incorrect responses; CRTs that deviated from individual means by more than 2.5 standard deviations and individuals who deviated from the respective group mean by more than three standard deviations were removed.

The experiment was designed to test the prediction that serial updating in running span would demand greater cognitive resources compared with simple serial recall. The temporal characteristics of these resources demands within the trials were speculated without any strong hypotheses, and this exploratory approach was evident in the statistical analyses. At the task level, the difference between single and dual CRTs across the three memory tasks was examined using a 2x3 ANOVA with two factors: load (single versus dual) and memory task (running span, modified span, and simple span). At the trial level, the difference in dual CRTs between early and late list positions was examined across the three tasks. For this, CRTs were partitioned into early positions (one to four) and late positions (five and six) to test if the effect of updating was found across the course of a running span list or specifically from items *n+1* onward with $n = 4$ in running span. CRTs associated with later items only included item five and six rather than all later updating items in running span. This was because these were the only updating items in running span that could be compared with items at the same positions in the other tasks, as simple span did not have later list positions. The CRTs at these positions were then analysed using a 2x3 ANOVA with two factors: position (early vs late) and memory task.

At the item level, the variation in dual CRTs during the interval between items was studied by dividing the inter-item interval of 2400 ms into six bins of 400 ms each, with two bins of stimulus presentation and four bins of silent post-presentation interval. These were then analysed using a 6x2x3 ANOVA with three factors: bin (six 400 ms bins), position (early versus late, separated at position four in the list as described above), and memory task. A similar analysis was also carried out to compare variation in dual CRTs at the item level specifically between updating items in running span (item 5 onwards) and modified span (items 8, 15, and 22).

Recall. Performance in the memory tasks was scored as the proportion of items recalled in their correct serial position. The data were screened for outliers deviating by more than three standard deviations from the group mean (none detected). The effects of load and target position on recall accuracy were investigated separately for each task. For this, three ANOVAs were used to test if there was a difference in recall accuracy between single and dual load conditions across the target positions (four in running span, seven in both simple span and modified span).

Previously, studies have investigated the detailed nature of the errors made in simple serial recall paradigms (Henson, Norris, Page, & Baddeley, 1996). Errors typically fall into the following three categories. *Transpositions* are the items presented in the list that were retrieved at incorrect positions, for instance when the item presented in the second position is recalled in the third output position. *Intrusions* are items that are retrieved but were not presented in the list, for instance by recalling an item from the previous trial. *Omissions* are those items that are not recalled because participants skipping retrieval at a particular position to elected to say ‘blank’ or ‘space’. The present study examined the error patterns in running span for the first time. This error analysis was exploratory and provided the opportunity to compare across standard serial recall and running span.

Post hoc tests were used to explore significant interactions terms for both RT and recall analyses. Bonferroni correction for multiple comparisons was applied, and both corrected and uncorrected p values are presented. A Greenhouse-Geisser correction was applied in case sphericity was violated in factors with repeated measures.

2.4 Results

The three groups completing running span, modified span and simple span were matched on age, $F(2,87) = 2.14$, $p = .12$, gender, $\chi^2(2, N = 90) = 0.12$, $p = .94$, and verbal short-term memory capacity measured using the digit span task, $F(2,87) = 0.21$, $p = .81$. Table 2.1 summarises participant characteristics, as well as mean RTs and recall accuracy, as a function of three memory tasks and attentional load (single and dual load).

2.4.1 Reaction times

Two participants were removed after outlier screening as outlined in Section 2.3.8, leaving 29 participants each in running span and simple span and 30 in modified span. The data from the CRT task are summarised in Figure 2.3 (across trials) and Figure 2.4 (across bins at the item-level).

2.4.1.1 Task-level analyses

CRTs in the single and dual load conditions were compared across the three groups completing different memory tasks using a 2x3 ANOVA. There was a main effect of load, such that responses in the single load condition were faster than in the dual load condition, $F(1,85) = 72.84$, $p < .001$, $\eta_p^2 = .46$. There was also a significant interaction between load and memory task, $F(2,85) = 5.20$, $p = .007$, $\eta_p^2 = .11$.

The interaction between load and task was explored using post hoc tests (Table 2.2), which showed that single CRTs did not differ between tasks. In contrast, dual CRTs in simple span were different from both running span, $t(56) = 2.22$, $p = .03$, and modified span, $t(57) = 2.07$, $p = .04$, but dual CRTs in running span and modified span were not significantly different, $t(57) = .47$, $p = .64$.

2.4.1.2 Trial-level analyses

Dual CRTs during early positions (one to four) and late positions (five and six) were compared across the three memory tasks using a 2x3 ANOVA. There was a significant effect of position, $F(1,85) = 56.22$, $p < .001$, $\eta_p^2 = .40$, as well as a significant interaction between position and task, $F(2,85) = 7.60$, $p = .001$, $\eta_p^2 = .15$. Post hoc tests presented in Table 2.3 showed that CRTs during early positions were not significantly different between any pair of memory tasks. The CRTs during the late positions in running span were significantly different from simple span, $t(56) = 2.57$, $p = .01$, while those in modified span were not significantly different from either simple span or running span, *p*s > .05.

2.4.1.3 Item-level analyses

CRTs across the six bins of the item interval (400ms each) were compared between early and late positions and across three memory tasks in a 6x2x3 ANOVA. Figure 2.4 summarises the RT data, and Table 2.4 presents the results from the omnibus ANOVA. There was a significant main effect of bin and position. All the two-way and three-way interactions were significant as well.

There was a significant three-way interaction between bin, position and task, $F(7.35,308.2) = 2.56$, $p = .01$, $\eta_p^2 = .06$. Two post hoc 6x3 mixed measure ANOVAs were conducted to explore the interaction between bin and task separately for early and late positions. During early positions, there was no interaction between bin and task, $F(6.0,255.1) = 1.11$, $p = .36$, $\eta_p^2 = .03$, while during late positions, a significant interaction was found, $F(6.3, 269.6) = 3.03$, $p = .006$, $\eta_p^2 = .07$.

Further post hoc tests showed that during these late positions, there was an increase in CRT from the second to the third bin (centred around 1000 ms) for all three tasks (Table 2.5). The magnitude of this peak was greatest in running span compared with modified span and simple span. Also, this CRT peak at 1000 ms was specific to update items in running span and was not found during update items in modified span (Table 2.6). In fact, in modified span there was a CRT dip at the same time-point, as also shown in Figure 2.4.

Table 2.1

Participant characteristics and mean ± SDs for performance in choice reaction time (CRT) task and memory task, for each load and task condition in Experiment 1.

		Running Span	Modified Span	Simple Span
Age (years)		25.47 ± 4.04	23.33 ± 3.77	24.33 ± 4.16
Gender		23 f, 7 m	22 f, 8 m	23 f, 7 m
Digit span		7.03 ± 1.33	7.17 ± 1.84	7.30 ± 1.56
CRT (ms)	Single	425 ± 49	427 ± 55	424 ± 43
	Dual	489 ± 90	480 ± 69	447 ± 51
Recall accuracy[1]	Single	.83 ± .10	.70 ± .21	.77 ± .15
	Dual	.76 ± .13	.60 ± .24	.67 ± .16

[1] Recall scored as proportion of items recalled in correct serial position

Table 2.2

Post-hoc independent sample t-tests of CRTs between each pair of tasks, computed for single and dual CRTs separately in Experiment 1.

	Running span versus Modified span				Running span versus Simple span				Modified span versus Simple span			
	Mean diff (ms)	*t* *(57)*	*p*	*d*	Mean diff (ms)	*t* *(56)*	*p*	*d*	Mean diff (ms)	*t* *(57)*	*p*	*d*
Single CRTs	3	.19	.85	.05	<1	.04	.97	.01	3	.24	.81	.06
Dual CRTs	10	.47	.64	.12	43	2.22	**.03**	.60	33	2.07	**.04**	.55

Note: The data analysed here include all CRTs after trimming and outlier correction, across all lists within each task (for dual condition). Bold text denotes significant effects at the p < .05 level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method.

Table 2.3

Post-hoc independent sample t-tests of dual CRTs between each pair of tasks analysed separately for early positions (1 to 4) and late positions (5 and 6) in Experiment 1.

	Running span versus Modified span				Running span versus Simple span				Modified span versus Simple span			
	Mean diff (ms)	*t* *(57)*	*p*	*d*	Mean diff (ms)	*t* *(56)*	*p*	*d*	Mean diff (ms)	*t* *(57)*	*p*	*d*
Early positions	3	.20	.84	.05	22	1.40	.17	.37	19	1.30	.20	.34
Late positions	61	1.89	.06	.52	80	2.57	***.01***	.74	19	1.04	.30	.27

Note: The data analysed here include dual CRTs, across the first six positions as these were comparable across the three tasks. Bold text denotes significant effects at the $p < .05$ level; bold italicised text indicates significance effects after adjusting for multiple comparisons using the Bonferroni method.

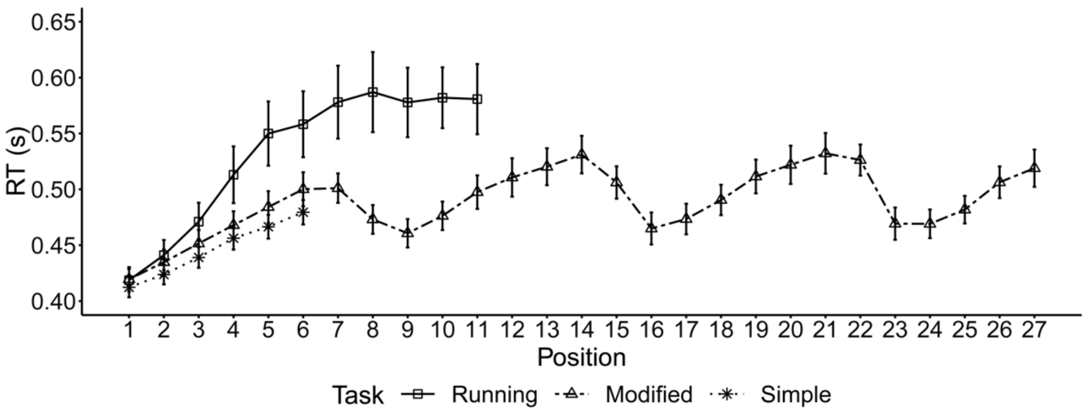

Figure 2.3 Mean concurrent CRTs across list positions for each memory task in Experiment 1. Note that the data are averaged across all list lengths, thus later positions in running and modified span contribute fewer data points. RTs associated with the final position across lists are not displayed here, see text for data exclusion. Error bars represent standard error of the mean.

Table 2.4

Results from the omnibus 6x2x3 ANOVA to investigate the effect of bin, position, and task in Experiment 1.

	F	df	p	η_p^2
Bin	8.81	3.25,6.50	**< .001**	.09
Position	64.54	1,85	**< .001**	.43
Task	2.96	2,85	.057	.07
Bin * Position	27.30	3.63,308.24	**< .001**	.24
Bin * Task	2.73	6.50,276.41	**.01**	.06
Position * Task	6.60	2,85	**.002**	.13
Bin * Position * Task	2.56	7.25,308.24	**.01**	.06

Note: Bin was a within-subject factor with six levels (the 2400 ms interval between consecutive items was divided into six 400 ms bins). Position was a within-subject factor with two levels (early positions 1 to 4 versus late positions 5 and 6). Task was a between-group factor with three levels (running span, simple span, and modified span). The data analysed here include dual CRTs after trimming and outlier correction extracted from the first six positions as those were comparable across tasks. Bold text denotes significant effects at the $p < .05$ level.

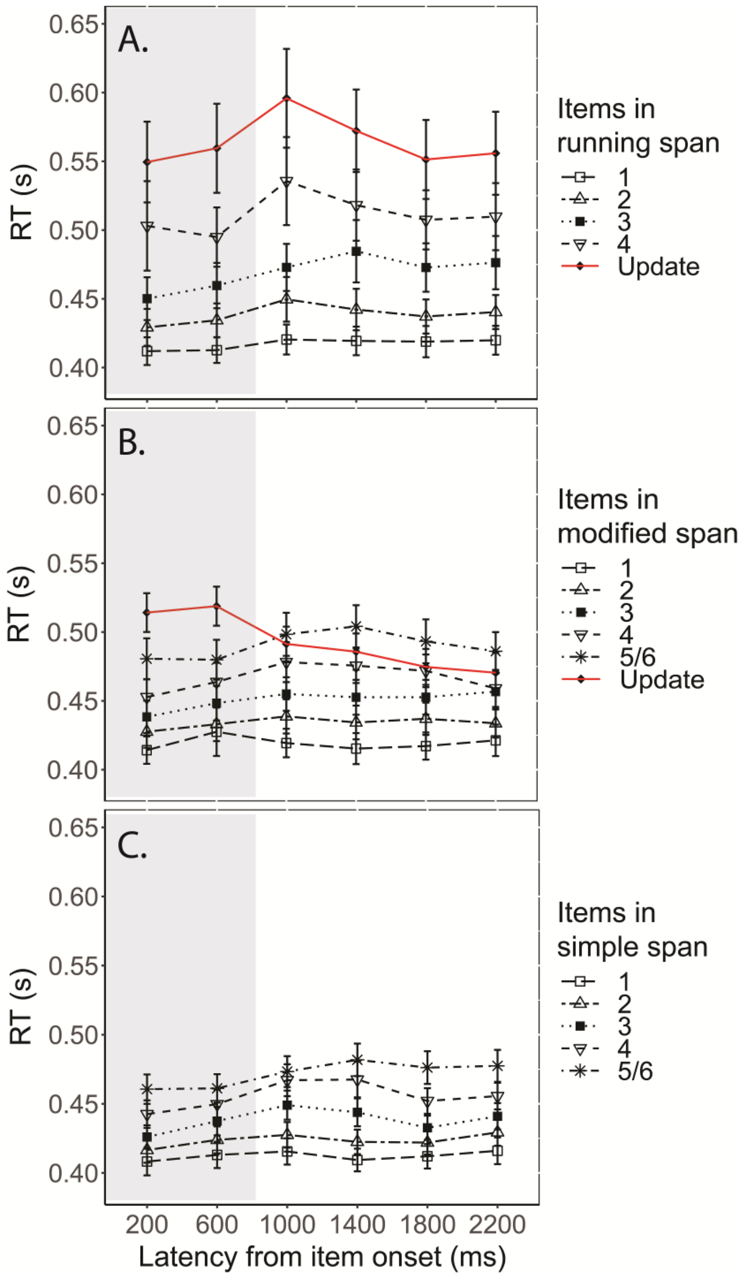

Figure 2.4 Mean concurrent CRTs as a function of latency from onset of memory item, plotted for separately across tasks in Experiment 1. Please note that while the data are illustrated per position, the analysis collapsed items into early and late positions (see text for more). The first 800 ms represent the duration of the item presentation (shaded in grey), followed by a 1600 ms silent inter-item interval (unshaded). (A) CRTs within running span, with position 1, 2, 3, 4 (black lines) and update items from position 5 onwards (red line). (B) CRTs within modified span, with position 1, 2, 3, 4, and 5/6 (black lines) and update items at positions 8, 15, and 22 (red line). (C) CRTs within simple span, with position 1, 2, 3, 4, and 5/6 (black lines) with no update items in the task. Error bars represent standard error of the mean.

Table 2.5

Post-hoc 2x3 ANOVAs to investigate the interaction between bin and task for each pair of consecutive bins across memory tasks, and the difference between means where applicable in Experiment 1.

	Task * Bin interaction effect				Mean difference between bins (ms)		
	F	*df*	*p*	η_p^2	Running	Modified	Simple
Bin 1 vs 2	1.46	2,85	.24	.03	.	.	.
Bin 2 vs 3	3.76	2,85	**.03**	.08	***40***	***18***	***12***
Bin 3 vs 4	3.95	2,85	**.02**	.09	- 21	6	**8**
Bin 4 vs 5	1.84	2,85	.17	.04	.	.	.
Bin 5 vs 6	1.41	2,85	.25	.03	.	.	.

Note: The data analysed here include dual CRTs at late positions (five and six). Pairwise analyses were conducted only if the interaction between task and bin for that bin-pair was significant. An increase in RT between consecutive bins is indicated in positive mean difference, while a decrease is indicated in negative values. Bold text denotes significant effects at the $p < .05$ level; bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni method. Only data related to late positions provided here as there was no overall bin*task interaction during in the 6x3 ANOVA conducted for early positions (see text).

Table 2.6

Post-hoc 2x2 ANOVAs to investigate the interaction between bin and task for each pair of consecutive bins between the two WM tasks involving updating (running span and modified span), and the difference between means where applicable in Experiment 1.

	Interaction between task and bin				Mean difference	
	F	*df*	*p*	η_p^2	Running	Modified
Bin 1 and 2	.18	1,57	.67	.003	.	.
Bin 2 and 3	25.62	1,57	***< .001***	.31	***36***	***- 27***
Bin 3 and 4	2.54	1,57	.12	.04	.	.
Bin 4 and 5	.81	1,57	.37	.01	.	.
Bin 5 and 6	1.29	1,57	.26	.02	.	.

Note: The data analysed here include dual CRTs from update positions for two memory tasks (running span: positions 5-12; modified span: positions: 8, 15, and 22 in the sequence). Pairwise analyses were conducted only if the interaction between task and bin for that bin-pair was significant. An increase in RT between consecutive bins is indicated in positive mean difference, while a decrease is indicated in negative values. Bold text denotes significant effects at the $p < .05$ level; bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni method.

2.4.2 Recall

2.4.2.1 Recall accuracy

The recall accuracy data from the memory tasks are summarised in Figure 2.5. Performance under single and dual loads was compared across target positions separately for each task using repeated measures ANOVAs. Post hoc paired-sample t-tests are presented in Table 2.7.

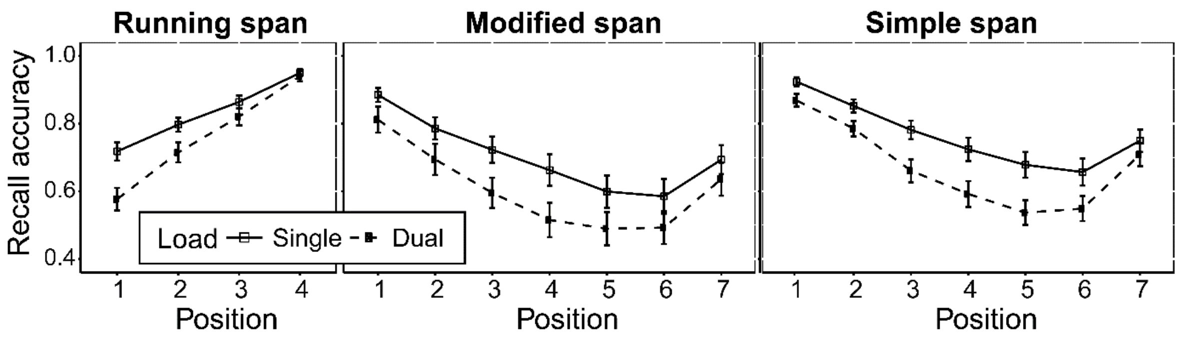

Figure 2.5 Mean recall accuracy (proportion of items recalled in correct serial position) across serial positions for the three memory tasks in single (solid line) and dual (dashed line) load conditions in Experiment 1. Error bars represent standard error.

In running span, there was a main effect of load, $F(1,29) = 22.19$, $p < .001$, $\eta_p^2 = 0.43$, and target position, $F(1.5,44.4) = 121.23$, $p < .001$, $\eta_p^2 = 0.80$. There was also a significant interaction between load and position, $F(2.3,66.4) = 19.68$, $p < .001$, $\eta_p^2 = 0.40$, with the effect of load decreasing across successive recall positions.

Recall in modified span exhibited a significant effect of both load, $F(1,29) = 47.59$, $p < .001$, $\eta_p^2 = 0.62$, and target position, $F(2.7,76.9) = 53.82$, $p < .001$, $\eta_p^2 = 0.65$. Load and position also showed a significant interaction, $F(3.3,94.4) = 3.55$, $p = .002$, $\eta_p^2 = 0.11$, such that the effect of load on recall increased from serial position one to four, and then decreased from position five onward.

Similarly, recall in simple span showed a significant effect of load, $F(1,29) = 53.29$, $p < .001$, $\eta_p^2 = 0.65$, and target position, $F(2.2,62.4) = 57.03$, $p < .001$, $\eta_p^2 = 0.66$. Load and position also showed a significant interaction, $F(4.4,128.4) = 8.61$, $p < .001$, $\eta_p^2 = 0.23$, such that the effect of load on recall increased from serial position one to five and decreased from position six to seven.

2.4.2.2 Error patterns

Errors in recall were divided into transpositions, intrusions and omissions, as outlined in Section 2.3.8. The probabilities of these errors for each memory task are summarized in Figure 2.6 as a function of the output position. Across all tasks, transpositions were the largest single source of error in recall, followed by omissions particularly at the first target position. Relatively low levels of intrusion were observed in the data.

The pattern of errors in modified span was similar to that in simple span, but these profiles were different from that in running span. In particular, the probability of transpositions at the first output position was the highest in running span compared with simple span and modified span. This observation may reflect the interference from the pre-target items in running span that were absent in simple span. The items in previous lists within the same trial in modified span did not appear to cause substantial interference as the errors in modified span and simple span were almost identical. In all WM tasks, a difference between the single and dual load conditions was observed in terms of the overall error levels, but the pattern of errors was similar between the two conditions.

The transposition gradients are presented in Figure 2.7, collapsed for single and dual loads in each memory task. At each output position, the peak represents the recall probability of the correct item flanked by the inaccurate recall of items presented at other positions. The gradients reflected the overall serial position curves in each task and transpositions appeared mostly symmetric around the peak.

Table 2.7

Post-hoc paired-sample t-tests of recall accuracy in single and dual load condition, performed separately for each target position and memory task in Experiment 1.

	Running span			Modified span			Simple span		
	Recall in single versus dual-task			Recall in single versus dual-task			Recall in single versus dual-task		
	t	*p*	*Cohen's d*	*t*	*p*	*Cohen's d*	*t*	*p*	*Cohen's d*
Position 1	6.17	***<.001***	1.16	3.27	***.003***	.89	3.87	***.001***	.75
Position 2	3.89	***.001***	.76	3.83	***.001***	.83	4.01	***<.001***	.76
Position 3	2.61	**. 014**	.52	7.27	***<.001***	1.41	6.04	***<.001***	1.15
Position 4	.99	.330	.19	7.54	***<.001***	1.41	5.78	***<.001***	1.07
Position 5	.	.	.	5.36	***<.001***	.98	7.04	***<.001***	1.28
Position 6	.	.	.	4.45	***<.001***	.88	7.02	***<.001***	1.33
Position 7	.	.	.	2.66	**. 013**	.44	2.5	**.018**	.45

Note: Bold text denotes significant effects at the $p < .05$ level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method.

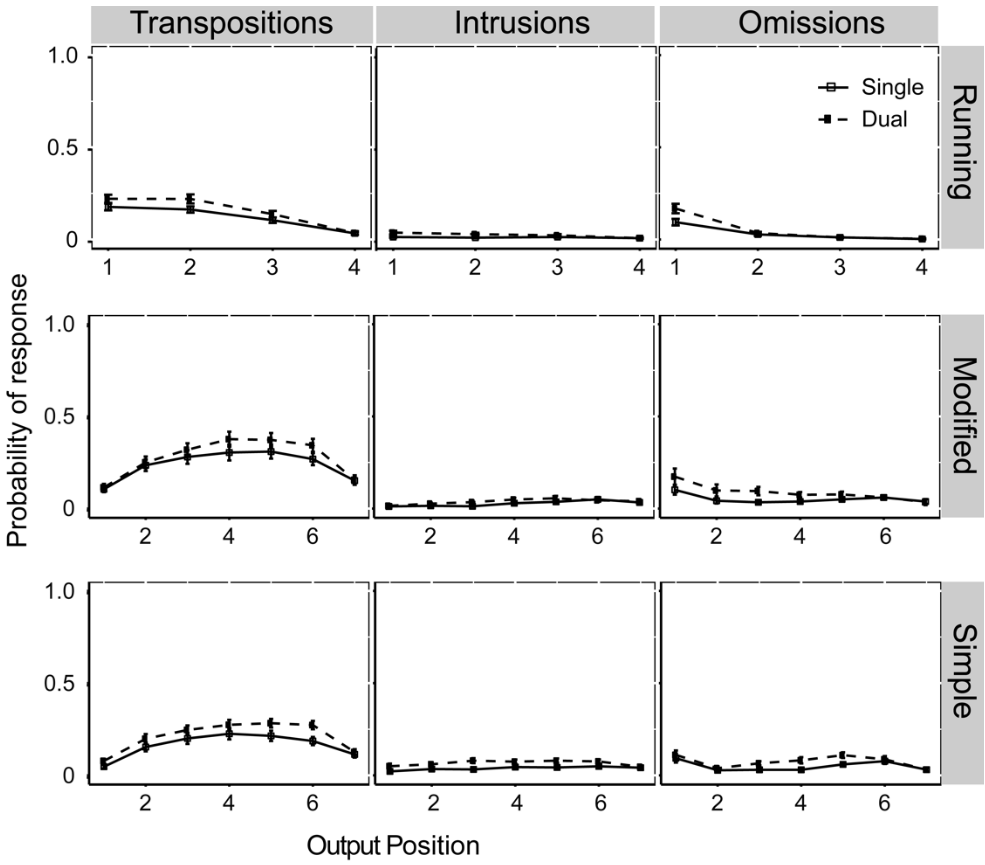

Figure 2.6 Recall error divided into transpositions (left), item intrusions (center) and omissions (right) at each recall position in the running span (top), modified span(middle) and simple span (bottom), presented separately for single load (solid line) and dual load (dashed line) conditions in Experiment 1.

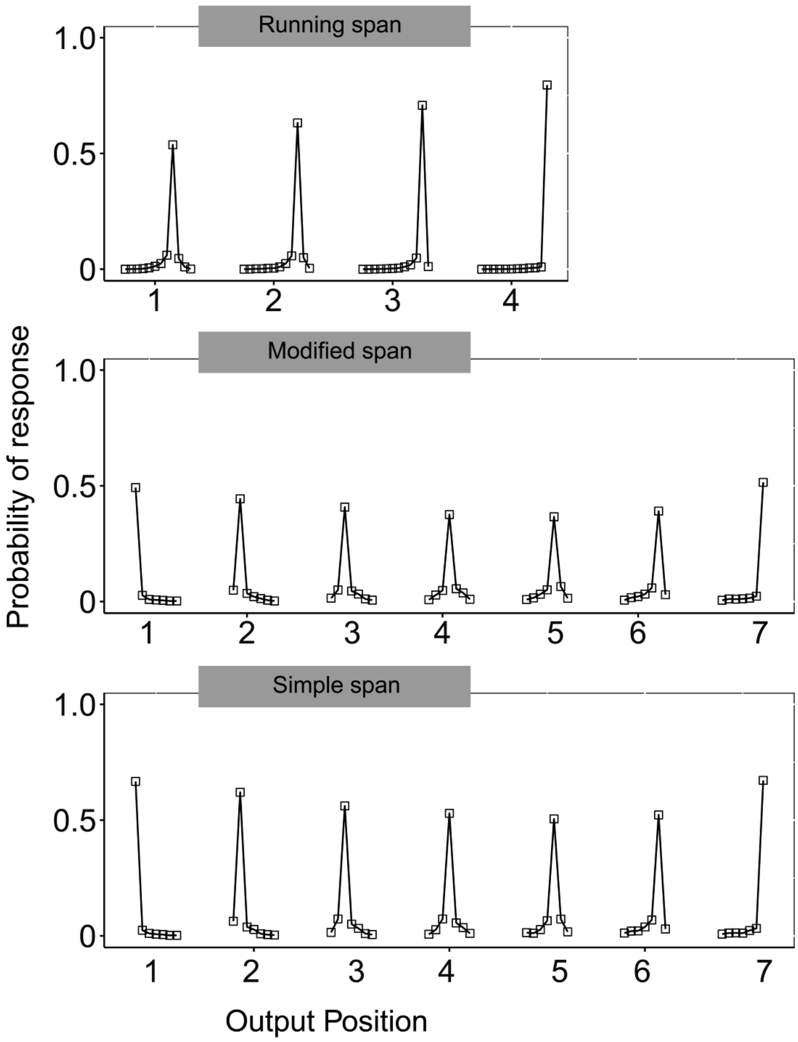

Figure 2.7 Transposition gradients for all three memory tasks in Experiment 1. At each output position on the horizontal axis, the peaks represent the likelihood of correct recall of input items, flanked by the likelihood of incorrect recall of items from neighbouring positions in the same encoded sequence. Note that in modified span, only the target list is presented, as the likelihood of error from non-target sequences presented earlier in the trials was negligible.

2.5 Discussion

The aim of Experiment 1 was to understand the process of serial updating using running span as an experimental vehicle of investigation. The experiment employed a divided attention approach, in which running span was performed along with a choice-reaction time task (CRT), and the CRTs served to indicate the resource demands associated with running span. Resource demands were compared with those incurred during two additional memory tasks, simple span and modified span. These tasks required serial recall in the absence of continuously keeping track of the latest set of target items. In this way, the resource demands specific to running span were identified and characterised over the course of the task.

The CRTs in the concurrent task were slower during running span than simple span. The relative increase in CRTs in running span was found after the *n*th position in the sequence. This suggests that more complex processes are involved in running span after *n* items are already stored in WM. At this point, simple encoding and maintenance may no longer suffice, and it may be necessary to trigger another process to keep track of relevant items in WM while discarding the earlier ones. The characterisation of a unique running span mode that is invariantly applied through the entire running span trial is inconsistent with these data. These findings favour instead the hypothesis that there is a shift in cognitive mode from maintenance to updating around position *n* and the demanding mode of updating is specifically applied when the list length is greater than the number of items to be recalled. The specificity of this updating-related demand to later positions is in line with previous proposals (Morris & Jones, 1990; Postle, 2003; Postle et al., 2001).

The general elevation in CRTs during the later positions in running span trials rose to a peak approximately 1000 ms from the onset of update items. This demand returned to the baseline established due to the updating mode before the onset of the next item in the memory list. This finding suggests that running span relies on a specific, time-bound process, additional to a general updating mode, in order to keep track of changes in the target recall set when prompted by the presentation of a new item. This observation supports previous suggestions that updating proceeds independently for each to-be-updated item and that these individual updating episodes do not cumulatively demand cognitive resources (Morris & Jones, 1990; Postle, 2003; Postle et al., 2001). Of the four alternatives outlined in the introduction, the data appeared to best fit the combination hypothesis (Figure 2.1).

A smaller CRT cost at the same time point of approximately 1000 ms from item onset was also evident at similar positions in the comparison tasks that did not require serial updating. This may reflect the demands imposed by general memory maintenance processes such as rehearsal or attentional refreshing likely to be common to all serial recall tasks (Barrouillet et al., 2004, 2009; Towse et al., 1998). The relative difference between running span and the other tasks is attributed to the additional process of continuous, serial updating required only in the former. It should be noted though that this differential demand function for running span cannot be simply explained in terms of task difficulty as recall accuracy in the single load condition was comparable for all three memory tasks.

The heightened CRT cost observed with a latency of around 1000 ms in running span could thus be a hallmark for the specific process of updating. From the data obtained in this experiment, it remains unclear if this time course is fixed or it aligns flexibly with the temporal conditions of the task. The updating process may be sensitive to the duration of item presentation or determined by the length of the inter-item interval. Alternatively, it may be it is time-locked such that, once activated, updating proceeds independent of the trial structure. Further investigation would be required to test these alternative accounts and fully understand the nature of updating.

A primacy effect was absent in running span but was evident in the two serial recall tasks. This lack of the primacy component is an established characteristic of the running span task (Bunting et al., 2006; Elosúa & Ruiz, 2008; Hockey, 1973; Morris & Jones, 1990; Postle et al., 2001; Ruiz & Elosúa, 2013a; Ruiz et al., 2005). Ruiz and colleagues have previously argued that this feature indicates the lack of active updating (Elosúa & Ruiz, 2008; Ruiz & Elosúa, 2013a; Ruiz et al., 2005). They suggest instead that participants passively listen to incoming stimuli until the end of the list and attempt to retrieve as many items as possible from sensory memory. This passive listening account however does not account for the CRT data in the present experiment. During early positions in the list, the resource demands in running span appeared to be similar to those in the other serial recall tasks suggesting the recruitment of active processes such as encoding and maintenance. During later positions at which the target updating would be expected, running span indeed showed heightened resource demands compared with the other tasks. Together, these observations suggested that the involvement of an active updating process that could not be captured by the recall data. This highlighted the utility of the divided attention approach and the index of resource demands provided by the concurrent CRT task.

The serial updating in running span was distinguished from the complete memory reset in modified span in terms of both CRT and recall data. A cyclical time course of the CRT costs was found in modified span trials. The demands increased with each new position in the 7-item target sequence (as in simple span). This returned to baseline once the encoding of a new target set was initialized and then ascended with each subsequent position in that set. The resource demands at the item level also showed considerable differences between modified span and running span. The update items in modified span did not show the characteristic elevation in CRT around 1000 ms found in update items in running span. In fact, the CRTs in modified span diminished, rather than peaked, in the interval following the reset items. This is in line with previous studies of item updating showing that a complete memory reset is rapid and while partial updates are time-consuming (Kessler & Meiran, 2008). These findings indicate that the process of restarting the encoding of entire new stimulus sequences imposes few cognitive demands and is largely equivalent to the process of starting at the beginning of a sequence. This conclusion is reinforced by the comparable serial position functions and error profile in modified span and simple span.

To conclude, this experiment demonstrated that running span involves a complex, demanding process. These data add temporal specificity to previous characterisations of the serial updating process (Bunting et al., 2006; Hockey, 1973; Morris & Jones, 1990; Postle, 2003; Postle et al., 2001) and could not be accounted by the passive listening account of running span (Elosúa & Ruiz, 2008; Ruiz et al., 2005). If the time course of resource demands presented above indeed reflects updating as proposed, then it should diminish under conditions when an active updating strategy cannot be adopted. In Experiment 2, this prediction was tested by manipulating the temporal parameters of the task and administering running span using fast presentation rates known to limit the application of updating (Bunting et al., 2006; Cowan et al., 2005; Hockey, 1973).

Chapter 3. Running span demands as a function of presentation rate

3.1 Overview

A series of exploratory analyses in Experiment 1 showed the running span task involves a complex, demanding process with a specific time-course. Importantly, this process was not apparent in similar serial recall tasks *without* working memory updating requirements. On this basis, it was proposed that the demand profiles observed in running span provide markers of the process of serial-updating. To test this proposal, in this chapter, Experiment 2 measured resource demands in a running span task under different task conditions, using the same divided attention approach employed in Experiment 1. It was hypothesised that the running span demands under conditions known to allow working memory updating would resemble those found in Experiment 1. Under task conditions not conducive to updating, it was anticipated that the updating-related demands would be absent.

Previously, it has been shown that the updating process in running span is sensitive to the rate of item presentation (Bunting et al., 2006; Hockey, 1973). Experiment 2 thus administered running span using different presentation rates. Consistent with the predictions, the running span demands previously observed in Experiment 1 were also found in Experiment 2 but only when the presentation rate was slower than one item per second. This suggested that the process underlying the demand function, arguably that of serial updating, was rate-sensitive and could only be applied when the interval was long enough between successive items. The time course of this updating process appeared inflexible despite varying temporal conditions. Recall and strategy data reinforced the conclusion that serial updating was limited to slow rates and participants used passive listening strategies at fast rates.

The results of Experiment 1 and 2 reported in this and the previous chapter have been accepted pending minor corrections as an article to the Journal of Experiment Psychology: Learning, Memory, and Cognition.

3.2 Introduction

Running span is a complex working memory (WM) task used to examine memory updating (Morris & Jones, 1990). In this task, participants are asked to recall the latest *n* items from long sequences of variable length. As sequences grow, items increasingly become outdated and need to be removed from WM while the latest, relevant items continue to be maintained. This process of WM updating proceeds on the basis of serial order. Not only does it require irrelevant material to be discarded but also the already encoded items have to be repositioned within the target set to accommodate newly presented information (see Section 1.3.2 for a detailed analysis of task demands).

While the precise mechanisms supporting the updating process are unclear, Experiment 1 characterised the time course of resource demands in running span. The experiment used a divided attention approach to show that demands in a trial were elevated at items located after position *n*, which is the point in the sequence at which updating could be reasonably expected to occur. The resource demands peaked 1000 ms after the onset of these items in running span compared with the demands recorded at similar list positions in other non-updating serial recall tasks. It was proposed that these demands reflect the serial-update process involved when participants are actively trying to keep track of the latest items in running span.

One way to test this serial updating proposal is to examine the resource demands in running span while manipulating the opportunity for participants to engage in an updating process. Previous research showed that an increase in the rate of stimulus presentation biases participants away from serial-updating and towards a less effortful passive listening strategy (Bunting et al., 2006; Hockey, 1973). Hockey showed that performance was impaired if participants had to actively keep up with relevant items when presented at a fast rate of 333 ms per item. This observation suggested that updating is a time-consuming process that cannot be optimally applied if stimulus presentation is too fast. Under rapid presentation conditions, performance was facilitated when participants were instructed to *not* pay special attention to the memory items and simply let them "float over [their] heads" until retrieval was cued (Hockey, 1973, p. 106). In other words, participants were explicitly discouraged from using any active strategies and asked to listen passively. Some researchers suggested that such a passive approach relies on the relatively automatic storage of incoming information, perhaps in the form of sensory memory traces, with little or no semantic processing (Broadway & Engle, 2010; Bunting et al., 2006; Cowan et al., 2005). Cowan et al. claimed that the undecayed sensory traces (of the most recent items) are immediately converted into a phonological or semantic form so that they can be recalled once the presentation is complete.

The rate of presentation used in running span can therefore be used to manipulate the strategy used to perform the task. Slower rates appear to allow for the application of a demanding serial-updating process while faster rates appear to discourage these strategies and lead participants to rely on a relatively effortless passive listening approach (Bunting et al., 2006; Hockey, 1973). However, the exact presentation rate associated with the anticipated switch in running span strategy is unclear. Previous investigations suggest that participants use a passive strategy at the presentation rates between 300 to 500 ms per item and shift to an active strategy at the rate of 1000 ms per item or slower (Botto et al., 2014; Broadway & Engle, 2010; Bunting et al., 2006; Collette et al., 2007; Cowan et al., 2005; Hockey, 1973; Kiss et al., 1998; Morris & Jones, 1990; Postle, 2003; Postle et al., 2001; Ruiz & Elosúa, 2013). Hockey examined the association between rate and strategy at three different presentation rates and found evidence for a continuous, gradual shift from one strategy to the other (Hockey, 1973; see also Hamilton & Hockey, 1974). This could indicate that participants have individual temporal limits associated with a shift in strategy but individual differences are blurred when data are aggregated at the group level. It could also be the case that at intermediate rates that are neither too fast nor too slow, such as 800 ms per item, participants can choose between the two strategies or combine them flexibly across the trials.

Experiment 2

Given the impact of presentation rate on the use of running span strategies, the primary aim of Experiment 2 was to examine the resource demands of running span administered using different rates of presentation. The running span demands were expected to decrease with an increase in the presentation rate, as a consequence of moving from an active strategy at slower rates to a passive strategy at faster rates. This experiment tracked the demands on cognitive resources using the same divided attention paradigm as in Experiment 1. A choice reaction time (CRT) task was administered at the same time as running span, and the variation in the CRTs was examined as a function of the presentation rate in the latter. If the CRT increase observed in the first experiment was indeed a marker of the serial-updating process, it should also be found in this experiment during the slow-paced running span task. There should be no increase in CRT when items are presented rapidly as participants would be expected to shift to an undemanding, passive listening strategy instead.

The secondary aim of this experiment was to examine the flexibility of the demand function associated with the updating process. Experiment 1 showed that the CRTs in the concurrent task were most elevated 1000 ms following the onset of update items. It was however unclear whether this finding reflected a time-locked cognitive process or if the properties of the task could determine the timing of the process. It may be that the peak is fixed, i.e. the complex operations underlying updating take 1000 ms from the onset of a new item to be activated and applied. If so, the CRT cost in this experiment would also be expected 1000 ms from item onset in conditions using intervals lengthier than 1000 ms between items. Alternatively, it may be that the increase in CRT in Experiment 1 emerged following the *offset* of items, each of which was presented for 800 ms. This cost might reflect, for example, the time taken to encode the new item before an update of the target set is initialised. To test this possibility, the duration of item presentation was decreased in Experiment 2 compared with that used in Experiment 1. A shorter item presentation ensured that item offset in Experiment 2 was earlier than that in Experiment 1. If it were the case that the timing of the update process was contingent on the offset of the update item, then the CRT elevation in this experiment would be evident earlier than in the previous experiment.

In summary, Experiment 2 tested if the heightened demands observed in the previous experiment reflected serial-updating. To this end, the rate of presentation in running span was varied to manipulate the opportunity to apply active strategies to perform the task. By changing the temporal conditions from the previous experiment, the experiment also investigated if the anticipated updating process was time-locked or flexibly deployed.

Pilot investigations

Two pilot investigations reported in Appendix C were conducted to determine the presentation rates that could demonstrate the difference between passive and active strategies in the task. Pilot study 1 tested running span behaviour at 500 ms/item and 800 ms/item. While the two conditions showed some differences in demand profiles, a clear effect of rate on resource demands was not apparent. To test the possibility that the manipulation was not strong enough to elicit the hypothesised rate effect, Pilot study 2 used a wider range in the rate of presentation. In this pilot study, the following three rate conditions were tested, 400 ms/item (fast rate), 800 ms/item (medium rate), and 1600 ms/item (slow rate). In each rate condition, the item was presented for 400 ms followed by a post-presentation interval of varying duration (none in fast, 400 ms in medium and 1200 ms in slow rate). There was clear evidence in Pilot 2 that the resource demands were sensitive to the rate of presentation, and the three rates were thus adopted for the main experiment, as described below.

3.3 Method

3.3.1 Participants

Thirty participants (12 male, 18 female, mean age = 24.3 years, *SD* = 3.9 years) who spoke English as a native language were recruited for the study. The eligibility criteria and ethical guidelines described in Chapter 2 were also applied for this study.

3.3.2 Procedure

A 3x2 design was used to investigate the effect of two factors, presentation rate and attentional load. Presentation rate was a within-subject factor comprised of three levels (fast, medium, and slow). Participants completed running span at all three presentation rates across sessions in a counterbalanced order. Attentional load was also a within-subject factor consisting of two levels. In each session, participants completed a single and a dual version of the respective running span variant.

Each participant provided informed consent to participate in the study and completed three sessions on separate days. Three tasks were administered in each session, a choice reaction time (CRT) task, a running span task (with presentation rates applied in a counterbalanced order across sessions), and a dual load condition in which the CRT task and running span were performed simultaneously. The order of the tasks within a block was fixed (CRT, running span, and dual-task). Each session started with a practice block that was followed by five experimental blocks. A strategy questionnaire was then administered before the conclusion of the session. Each session typically lasted for an hour.

3.3.3 Tasks

The task structure of running span in this experiment was different from that in Experiment 1 in two ways. First, this experiment used items with a shorter presentation duration. For this, a new stimulus set of spoken letters was recorded so that the presentation of the memory items (i.e. letters) lasted for 400 ms. A female speaker of native British English was recorded as she spoke the letters at a rate of two letters per second. The recorded audio was divided into distinct letters; using Adobe Audition 3.0, each sound file was adjusted and compressed into a 400 ms duration in line with the P-centre approach (Morton et al., 1976). Second, this experiment used three rates of presentation. Each condition presented items for 400 ms using the new stimuli as described above. The presentation rate was varied by changing the duration of the (silent) interval between successive items in a memory sequence. In the fast rate, there was no interval between items, such that lists were continuously presented at a rate of 400 ms/item. In the medium rate, item presentation consisted of alternating items (400 ms) with a silent interval (400 ms), such that the overall rate of presentation was 800 ms/item. In the slow rate, the duration of the inter-item interval was 1200 ms, such that the overall presentation rate was 1600 ms/item.

All other features of the running span task, the CRT task and the dual load condition were the same as in Experiment 1 (Section 2.3).

3.3.4 Strategy reports

At the end of each session, participants were provided with a list of six strategies adapted from Norris, Hall and Gathercole (2019). The strategies were: (a) Passively receive the letters, (b) Rehearse or repeat the letters as they were presented, (c) Keep up with the last four letters as they were presented, (d) Group the letters by separating them into sets of particular sizes, (e) Group the letters according to their meaning, and (f) Form a mental image of the letters. Participants were asked to rate the frequency with which they used each strategy on a four-point scale ranging from almost never (=0) to occasionally to frequently to almost always (=3). At the end of the final session, participants were also asked to report the rate condition they experienced as the least challenging.

3.3.5 Statistical power

An *a-priori* power analysis was conducted using G-Power (Faul, Erdfelder, Lang, & Buchner, 2007). For this, an effect size was computed based on the RTs measured in the second pilot between the three rate conditions (Appendix C). With effect size f = 1.14 (as determined within G-Power), alpha = .05 and *a-priori* power = .95, it was determined that a total sample size of fifteen would be sufficient to run this experiment. As this was determined based on overall differences between tasks rather than at the fine-grained difference observed at the trial- and item-levels, it could be that the effect size based on pilot data was an overestimate (Thabane et al., 2010). As a result, a more conservative sample size of thirty participants was employed. This was also consistent with the sample size per group used in Experiment 1 in which the post hoc power for all reported effects were greater than .90.

3.3.6 Analysis plan

Data were trimmed and screened for outliers as in the first experiment.

Reaction time. Experiment 2 was designed to test hypotheses based on the results of Experiment 1 and thus used a confirmatory analysis approach to examine differences between concurrent RTs at the fast and slow rates. At the task level, a 2x3 ANOVA was used to examine the variation in concurrent RTs as a function of two factors: attentional load (single versus dual CRTs) and rate of presentation (fast, medium and slow rates). An interaction between load and rate was predicted, such that dual CRTs would be larger for the slow than fast rate condition.

At the trial level, a 2x3 ANOVA was used to examine the effect of list position (early items up to position four (= n) versus late items from position five in the list) and presentation rate on dual CRTs. An interaction between position and rate was expected, such that the late CRTs would be larger for the slow than fast rate condition.

At the item level, the interval between successive item onsets (800 ms in the medium rate and 1600 ms in the slow rate condition) was segmented into 400 ms bins. This division resulted in two bins at the medium rate and four bins at the slow rate. The item intervals in the fast rate condition could not be further divided as these were already 400 ms; this condition was thus not included in this analysis. In the medium rate condition, a 2x2 ANOVA was used to investigate concurrent RTs as a function of position (early versus late, as above) and bin (presentation versus post-presentation bin). In the slow rate condition, a 2x4 ANOVA similarly examined concurrent RTs as a function of position (early versus late) and bin (one presentation bin and three post-presentation bins). An interaction between task and bin was expected in the slow rate and planned comparisons were used to test the timing of the anticipated peak in RTs. Assuming that the updating-related peak is time-locked to 1000 ms from item onset, an RT increase was anticipated from the second to the third bin during the late positions in the slow rate condition (replicating the result from Experiment 1). If instead, the peak was found earlier in the interval, e.g. from first to the second bin, this would suggest that the updating process was rate-sensitive and flexible in time.

All analyses of CRTs in the medium rate condition were treated as exploratory as the relative use of an active strategy at this rate was unclear.

Recall. A 2x4x3 ANOVA was conducted to examine recall accuracy as a function of load, target position and rate of presentation. This tested if there was a difference between single and dual recall across the four target positions in the three rate conditions. Significant main effects of rate, position, and load were predicted, and post hoc tests were used to explore interaction effects.

Strategy use. Separate non-parametric Friedman tests were used to compare the effect of presentation rate on the mean ratings of frequency for the six listed strategies.

3.3 Results

Table 3.1 summarises findings for the CRT and running span tasks as a function of the rate of presentation (fast, medium, and slow) and attention load (single and dual load).

Table 3.1

Participant performance in concurrent CRT task and running span, for each load and rate condition in Experiment 2.

		Fast rate	Medium rate	Slow rate
RT (ms)	Single	399 ± 43	404 ± 51	403 ± 53
	Dual	384 ± 39	401 ± 45	418 ± 57
Recall accuracy[1]	Single	.79 ± .11	.80 ± .12	.81 ± .11
	Dual	.77 ± .13	.77 ± .14	.73 ± .13

[1] Recall scored as proportion of items recalled in correct serial position

3.3.1 Reaction time

Outlier screening based on the criteria as outlined in Section 2.3.8 resulted in the exclusion of one participant from the following analyses. The CRT data are summarised in Figure 3.1a (across trials) and Figure 3.1b (across the item-interval).

3.3.1.1 Task-level analysis

A 2x3 ANOVA compared single and dual CRTs across the three rate conditions. This showed a significant main effect of rate, $F(2,56) = 3.72$, $p = .03$, $\eta_p^2 = .12$, and a significant interaction between load and rate, $F(2,56) = 36.39$, $p < .001$, $\eta_p^2 = .57$. Planned comparisons (Table 3.2) showed that dual CRTs in the slow-paced running span were longer than in the fast-paced task as predicted. It was also found that dual CRTs in the medium rate were shorter than those in the slow rate but longer than those in the fast rate.

3.3.1.2 Trial-level analysis

A 2x3 ANOVA was conducted to compare dual CRTs at early and late positions, separated in a list at position four (= n) across the three presentation rates. Significant main effects were found for both position, $F(1,28) = 76.59$, $p < .001$, $\eta_p^2 = .73$, and rate, $F(2,56) = 8.69$, $p = .001$, $\eta_p^2 = .24$. There was also a significant interaction between position and rate, $F(1.6,45.9) = 13.44$, $p < .001$, $\eta_p^2 = .32$. In line with predictions, planned comparisons (Table 3.3) showed that the CRTs were greater in the slow than fast rate, and this difference was larger during the late than early positions in the sequence. Additionally, it was found that the CRTs in the medium rate condition were greater than those in the fast rate condition but only during late positions.

3.3.1.3 Item-level analysis

In the slow-paced running span, CRTs were examined across the four 400 ms bins across the item interval (one presentation and three post-presentation bins) and between early and late positions using a 4x2 ANOVA. Results showed significant main effects of both bin, $F(2.1,56.2) = 9.85$, $p < .001$, $\eta_p^2 = .26$, and position, $F(1,28) = 56.69$, $p < .001$, $\eta_p^2 = .68$. A significant interaction effect between the two factors was also found in line with predictions, $F(3,84) = 7.17$, $p < .001$, $\eta_p^2 = .20$. Planned comparisons (Table 3.4) showed that, during late list positions, the CRTs were stable from the first to the second bin, increased from the second to the third bin, and then decreased from the third to the fourth bin. Together, this indicated an RT peak approximately 1000 ms from the onset of items at late list positions in the slow rate condition. There was no significant variation in CRTs across bins in early positions in running span.

A further 2x2 ANOVA was conducted to examine CRTs between the two bins (one presentation and one post-presentation) during early and late positions in the medium rate condition. A significant main effect of position such that late items were associated with longer CRTs than early items, $F(1,28) = 83.92$, $p < .001$, $\eta_p^2 = .75$. The effect of bin was not significant, $F(1,28) = .001$, $p = .98$, $\eta_p^2 < .001$, and there was no significant interaction between bin and position, $F(1,28) = 1.24$, $p = .28$, $\eta_p^2 = .04$.

Table 3.2

Paired sample t-tests of CRTs between each pair of rate conditions separately for single and dual load condition in Experiment 2.

	Fast versus Medium				Fast versus Slow				Medium versus Slow			
	Mean Diff (ms)	*t* *(28)*	*p*	*d*	Mean Diff (ms)	*t* *(28)*	*p*	*d*	Mean Diff (ms)	*t* *(28)*	*p*	*d*
Single CRTs	5	.72	.48	.13	4	.58	.56	.11	<1	.08	.94	.01
Dual CRTs	17	2.74	***.01***	.51	33	4.08	***<.001***	.81	17	2.21	**.04**	.42

Note: The data analysed here include all CRTs after trimming and outlier correction, across all lists within each rate. Bold text denotes significant effects at the $p <$ .05 level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method. The mean difference reflects the mean RT of the faster rate subtracted from that of the slower rate for each respective comparison, e.g. medium RTs – fast RTs.

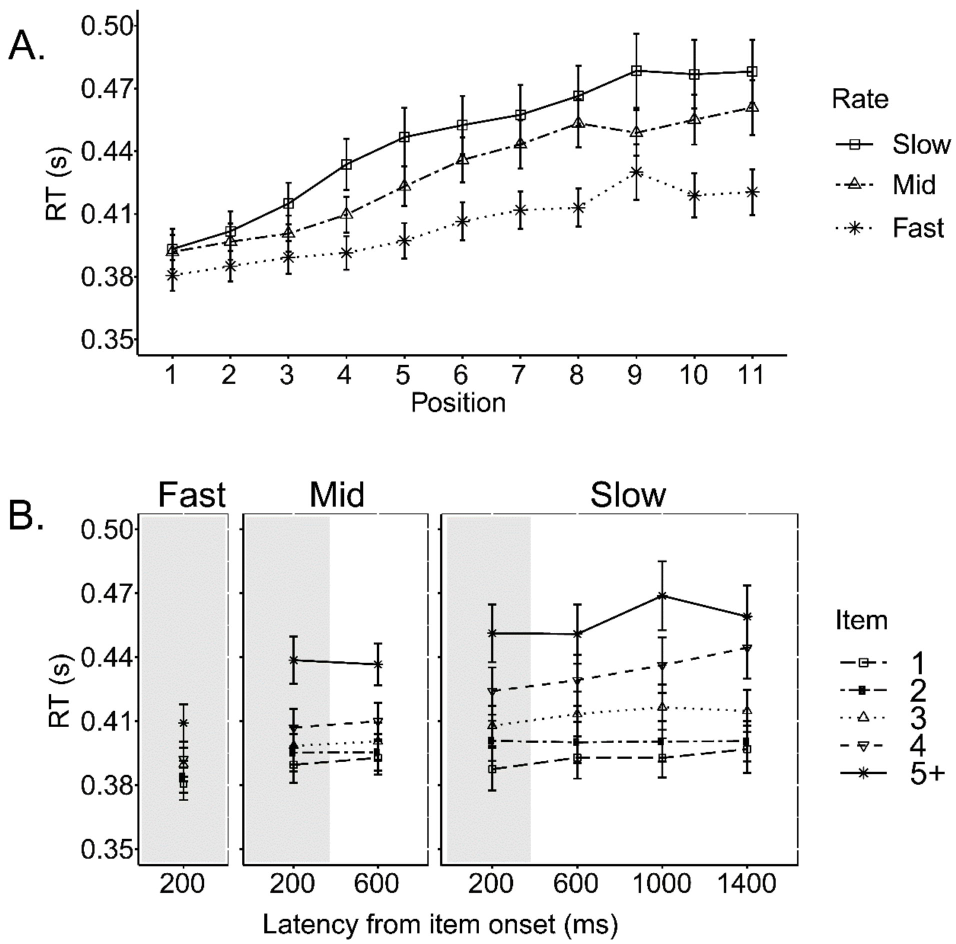

Figure 3.1 CRT data at the trial and item level in Experiment 2. (a) Mean concurrent CRTs across list positions for each rate condition. Note that the data are averaged across all list lengths; thus later positions contribute fewer data points. RTs associated with the final position across lists are not displayed here, see text for data exclusion. (b) Mean concurrent CRTs as a function of latency from item onset of memory item for all three rate conditions, separated by position 1, 2, 3, 4, and update items at position 5 onwards. Please note that while the data are illustrated per position, the analysis collapsed items into early and late positions (see text for more). The first 400 ms represents the duration of the item presentation (shaded in grey), followed by a variable duration of silent inter-item interval (unshaded). Error bars represent standard error of the mean.

Table 3.3

Paired sample t-tests of CRTs between each pair of rate conditions, separately for early and late positions in Experiment 2.

	Fast versus Medium				Fast versus Slow				Medium versus Slow			
	Mean Diff (ms)	*t* *(28)*	*p*	*d*	Mean Diff (ms)	*t* *(28)*	*p*	*d*	Mean Diff (ms)	*t* *(28)*	*p*	*d*
Early	11	1.83	.08	.35	22	3.03	***.005***	.31	11	1.71	.10	.32
Late	28	3.48	***.002***	.31	48	4.0	***<.001***	.31	20	1.78	.09	.31

Note: The data analysed here include all dual CRTs after trimming and outlier correction, across all lists within each rate. Early CRTs are those at positions one to four, and late CRTs are from positions five onward. Bold text denotes significant effects at the $p < .05$ level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method. The mean difference reflects the mean RT of the faster rate subtracted from that of the slower rate for each respective comparison, e.g. medium RTs – fast RTs.

Table 3.4

2x2 ANOVAs to examine the interaction between bin and position for each pair of consecutive bins between early and late positions within the slow rate condition, and pairwise mean differences in RT where applicable in Experiment 2.

	Interaction between Bin and Position				Pairwise mean diff in RT	
	F	*df*	*p*	η_p^2	Early	Late
Bin 1 vs 2	1.25	1,28	.27	.04	.	.
Bin 2 vs 3	17.66	1,28	***<.001***	.39	2	***18***
Bin 3 vs 4	8.99	1,28	***.006***	.24	2	***- 10***

Note: The data analysed here include dual CRTs in the slow rate condition after trimming and outlier correction extracted from early (one to four) and late positions (five onward). Pairwise analyses were conducted only if the interaction between bin and position for that bin-pair was significant. An increase in RT between consecutive bins is indicated in positive mean difference values, while a decrease is indicated in negative values. Bold text denotes significant effects at the $p < .05$ level, bold italicised text indicates significance effects after adjusting for multiple comparisons using the Bonferroni method.

3.3.2 Recall

3.3.2.1 Accurate recall

Figure 3.2 summarises recall performance in the single and dual load conditions, across the four target positions and the three presentation rates used in running span. To examine the effect of each of these factors, a 2x4x3 ANOVA was conducted. As predicted, there was a significant main effect of load, $F(1,28) = 24.29$, $p < .001$, $\eta_p^2 = .47$, and position, $F(1.3,34.7) = 126.84$, $p < .001$, $\eta_p^2 = .82$. Inconsistent with predictions, there was no significant effect of rate, $F(2,56) = 1.63$, $p = .21$, $\eta_p^2 = .06$. A three-way interaction between load, position and rate was found, $F(6,168) = 2.44$, $p = .03$, $\eta_p^2 = .08$, which was explored using post hoc tests (Table 3.5). In the slow rate condition, recall accuracy in single load was higher than dual load across all positions, with a greater dual-task impairment during the first two recall positions relative to the last two positions. There was no significant difference between the load conditions across all positions at the fast rate. In the medium rate, there was a trend suggesting a dual-task impairment difference at early positions, but not late target positions.

3.3.2.2 Errors in recall

Recall errors were divided into three categories as in Experiment 1. Briefly, transpositions included input items retrieved in wrong output positions, including pre-target items retrieved in target positions. Intrusions were items that were not present in the input sequence. Omissions were items for which retrieval was skipped, for example by saying "blank". Figure 3.3 summarises the probability of the three types of recall errors as a function of the output position in the target recall set, across the three rates and two loads. Across all conditions, transpositions were the most common errors followed by omissions, particularly at the first output position. Relatively low levels of intrusion were observed. There was little difference in recall error profiles between load and rate conditions, with one exception: in the slow task, there were more transpositions in the dual than single load condition. This difference between single versus dual load conditions was not observed at either the fast or medium paced tasks. The transposition gradients are presented in Figure 3.4. The data across the single and dual loads were collapsed, as there was a minimal effect of load on transpositions across the running span rate conditions. The gradients were mostly symmetric around the peak and reflected the overall serial position curves. The function appeared similar across the three rate conditions, suggesting that the transposition gradients were insensitive to the presentation rate.

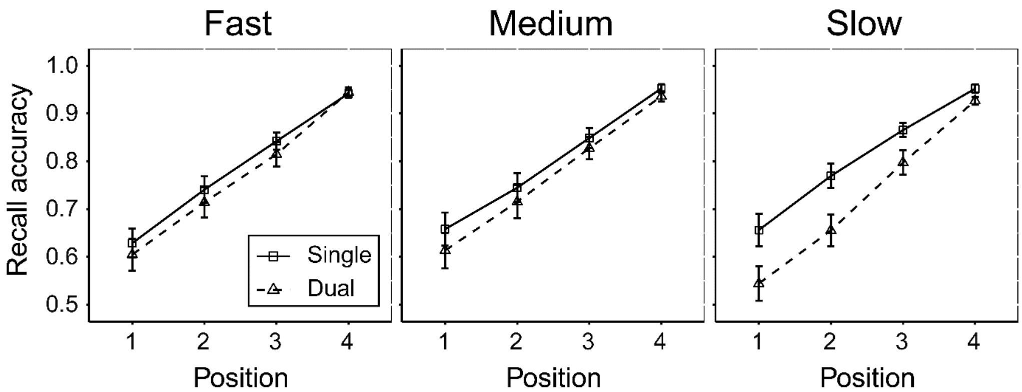

Figure 3.2 Mean recall accuracy (proportion of items recalled in correct serial position) across each output position for the three rate conditions across single (solid line) and dual (dashed line) load conditions in Experiment 2. Error bars represent standard error of the mean.

Table 3.5

Post-hoc paired-sample t-tests of recall accuracy in single and dual load performed separately for each target position and rate condition in Experiment 2.

	Fast rate			Medium rate			Slow rate		
	Recall in single versus dual load			Recall in single versus dual load			Recall in single versus dual load		
	t	*p*	*Cohen's d*	*t*	*p*	*Cohen's d*	*t*	*p*	*Cohen's d*
Position 1	1.46	.16	.28	2.60	**.02**	.49	5.61	***<.001***	1.04
Position 2	1.56	.13	.29	2.04	**.05**	.40	5.75	***<.001***	1.13
Position 3	1.74	.09	.36	1.55	.13	.29	2.97	***.006***	.62
Position 4	.27	.78	.05	1.68	.10	.31	3.18	***.004***	.59

Note: Bold text denotes significant effects at the $p < .05$ level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method.

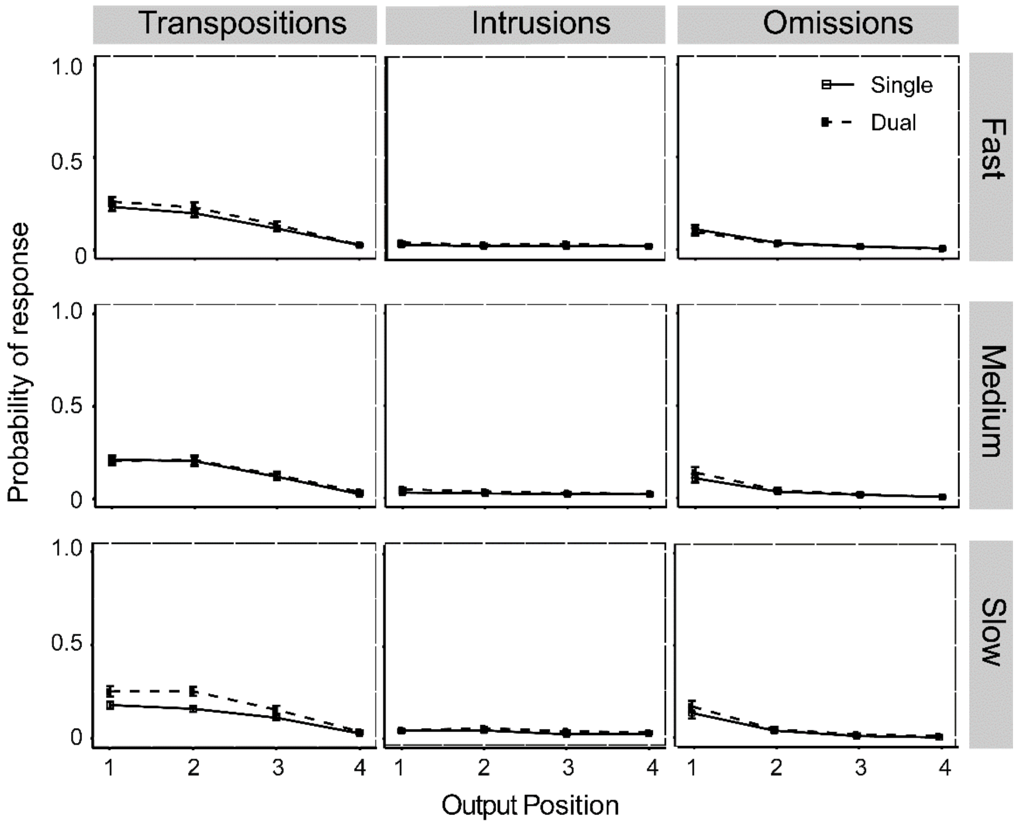

Figure 3.3 Mean recall errors in Experiment 2. Divided into transpositions (left), item intrusions (center) and omissions (right) at each recall position in the fast (top), medium (middle) and slow rate conditions (bottom), presented separately for single load (solid line) and dual load (dashed line) conditions.

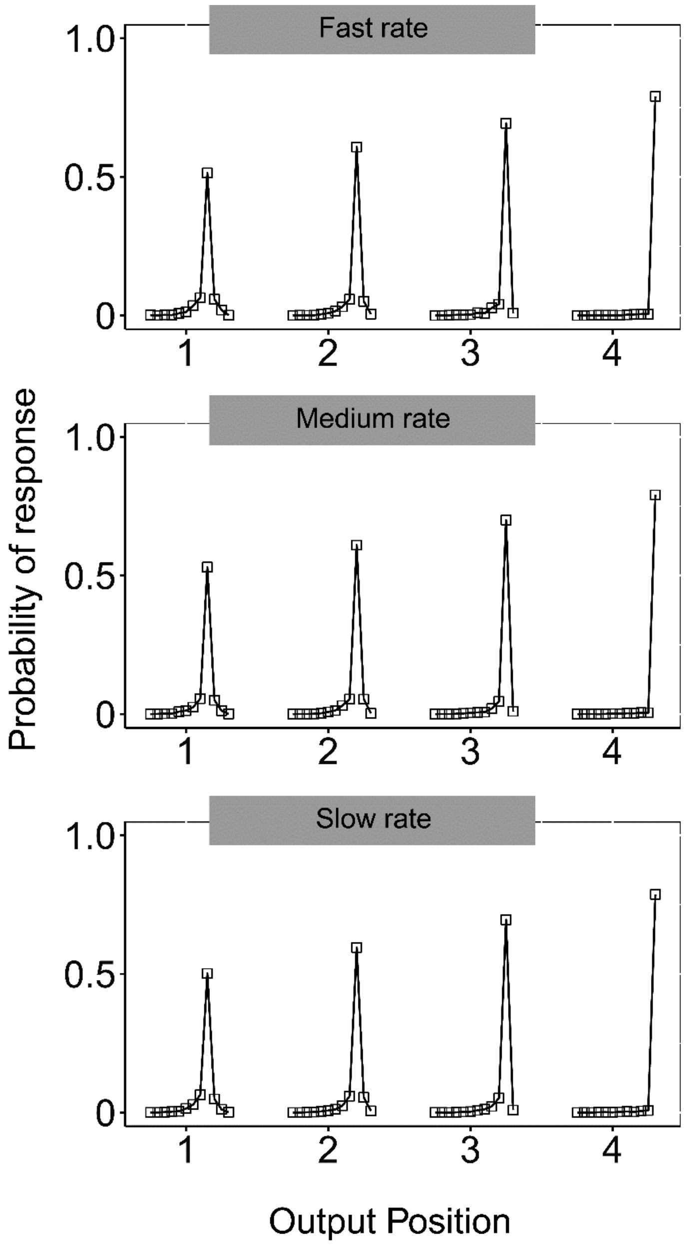

Figure 3.4 Transposition gradients for all three rate conditions in Experiment 2. At each output position on the horizontal axis, the peaks represent the likelihood of correct recall of input items, flanked by the likelihood of incorrect recall of items from neighbouring positions in the same encoded sequence.

3.3.3 Strategy use

Significantly more participants reported the running span tasks with the fast rate (N = 15) and medium rate (N = 12) as less challenging than the task with the slow rate (N = 3), $\chi 2$ (2, N = 30) = 7.8, p < .05. The reported frequency of use of six possible strategies is shown in Figure 3.5. Overall, participants reported employing fewer strategies in the fast-paced task (0.93) than the medium-paced (mean frequency rating = 1.13) or the slow-paced task (mean frequency rating = 1.39). The five strategies associated with an active approach in running span were most frequently employed in the slow rate, followed by the medium and fast rate conditions respectively. The passive listening strategy was the only exception to this trend as it was most frequently reported in the fast-paced task. Non-parametric Friedman tests comparing the self-reported use across the three rates showed that frequency ratings were significantly different for all strategies, all *p*s < .05, except for visualisation.

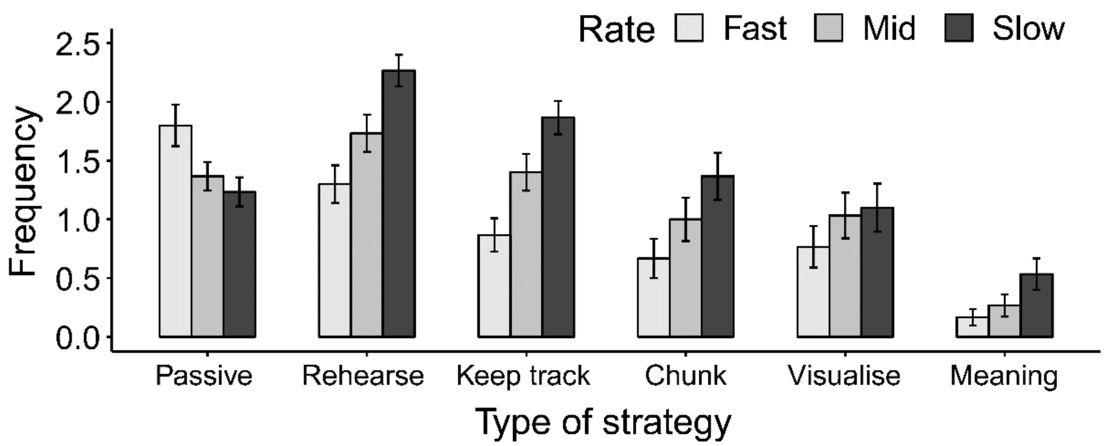

Figure 3.5 Mean frequency of self-reported strategy use for each rate condition in Experiment 2. Participants rated their use of each strategy from 0 (=almost never) to 3 (=almost always), see text for strategy statements.

3.4 Discussion

This experiment contrasted the resource demands imposed by three running span tasks differing in their rate of item presentation. The demands on cognitive resources were indexed using the CRTs recorded in a concurrent task. In line with predictions, the concurrent CRTs were considerably longer during the slow than fast presentation of information in running span, while intermediate costs were observed in the medium rate condition. Although the CRTs were elevated over the entire course of the slow-paced trials, the difference between rate conditions was more pronounced at the updating positions in the sequence. This finding is consistent with the predicted increase in the use of active, resource-demanding strategies as the interval between successive items lengthened, i.e. the rate slowed. This account also fits with participant reports that active strategies were most frequently adopted at a slow pace while passive listening was most preferred at a fast pace. Together, these data extend previous findings that active and passive approaches to running span are optimal at different presentation rates (Bunting et al., 2006; Hockey, 1973). It is the first known demonstration of the association between presentation rate and strategy choice in a repeated-measures design, clearly indicating that participants did not persevere in using the same strategy across different sessions and were instead able to flexibly change their approach to accommodate the temporal conditions of each session.

An increased delay in CRTs was found at 1000 ms following the onset of items in positions *n* onwards, replicating the demand profiles observed in Experiment 1, despite the difference in presentation rate between the experiments (2400 ms vs 1600 ms per item). The invariance in the timing of this feature suggests a fixed time course for the updating process, rather than one determined by presentation rate or item duration. A comparable CRT cost was not found in the medium rate condition, indicating that the 800 ms between successive items was not long enough to allow updating. This clarifies why recall is impaired when participants try to update the target set at rates faster than 1000 ms per item (e.g., Hockey, 1973).

Unexpectedly, it was found that dual RTs were faster than single RTs in the fast-paced condition. The pilot data showed the same finding, suggesting it is a reliable characteristic of the present experimental design (Appendix C). Results might reflect an entrainment effect of the presentation rate. In other words, it may be that participants strategically responded in the concurrent CRT task in time with the beat generated in the running span task due to periodic item presentation.

Contrary to previous observations (Bunting et al., 2006; Hockey, 1973), recall indices were not sensitive to the rate of presentation. There was no difference across the different rate conditions in terms of overall recall accuracy, serial position functions, error rates or transposition gradients. It could be that the experimental manipulation was not strong enough to reveal differences in the recall data. This interpretation however is unlikely as the presentation rates used in this experiment were similar to those used in the previous studies that showed a rate effect on recall performance (e.g. Bunting et al., 2006; Hockey, 1973). Another possibility is that the size of the recall set used in the present study (= four) was too small to reveal the anticipated rate effect. Bunting et al. found that a slow presentation rate improved recall particularly at positions four to six from the end of the list. In comparison, there was little impact of rate at positions closer to the end of the list. This observation suggests that the active and passive strategies employed at the slow and fast rates respectively are equally effective for the most recent items. As a result, the differential impact of presentation rate may only be apparent when bigger target sets are used.

One feature of recall performance that did vary with the presentation rate was the dual-task cost. There was a selective impairment in recall performance during the dual load condition in the slow-paced task. This dual-task cost was observed for the entire target set and appeared to decrease with each position. It was also observed in terms of the transposition errors in the recall data. The slow rate task showed higher transpositions, particularly at early output positions, in the dual than single-load condition. The division of resources between running span and the CRT task thus appeared to reduce recall accuracy disproportionately during slow item presentation. This finding is in line with the claim that an executively-mediated update process is explicitly applied only when there is sufficient time between items, recruiting resources away from, and thereby reducing performance in, the memory task (Bunting et al., 2006; Morris & Jones, 1990).

In summary, the divided attention approach has been used across two experiments to establish for the first time a consistent time course to the resource demands during a running span task. It was discovered that the elevation in resource demands 1000 ms after the presentation of an item at or beyond position *n* is absent when the tasks either do not require updating (Experiment 1) or preclude the process in running span (Experiment 2). While this CRT-based methodology provides the benefit of precisely tracking the changing resource demands of the running span task, sole reliance on this method may be unwise. The risk of any dual-task procedure is that it may fundamentally change behaviour on the primary task of interest, thus inducing a “dual-task mode” rather than genuinely reflecting performance under standard conditions. It would therefore be essential to use an alternative method to test if the outcomes of the two studies indicate serial-updating or if they are influenced by the divided attention method. Experiment 3 addressed this limitation by using a self-paced, single-task procedure presented in the following chapter. This third experiment enabled cognitive processing to be studied directly from the running span task instead of a concurrent task. Another advantage of this method was that it allowed for the impact of strategy instruction on task performance to be studied, instead of indirectly inducing strategies using task conditions as in this experiment. It may also be that the results presented so far are specific to running span rather than reflecting a general updating process applicable across other tasks with similar task structures. Experiment 4 thus tested this possibility by examining the resource demands in an *n*-back task (Chapter 5).

Chapter 4. Examining cognitive strategies in running span

4.1 Overview

There is a strong association between the rate of presentation in running span and the cognitive strategy used to perform the task. Previous studies have used memory performance at the *end of the list* to demonstrate that the optimal strategy at faster rates is passive listening, while active updating is more effective at slower rates (Bunting et al., 2006; Hockey, 1973). Experiment 2 previously established that presentation rate influences a participant's strategy use in running span, inferring processing demands of each strategy from the performance in a *concurrent task*. Across these studies however, the association between strategy and presentation was not tested using indices obtained *during* the presentation phase *directly* from running span. This is critical as such indices may not capture the cognitive processing during the phase when the strategies are applied and may be contaminated by the presence of the concurrent task.

This chapter presents Experiment 3 that addressed these limitations by employing a converging operations approach. The research question from Experiment 2 was inverted in Experiment 3 and the aim was to determine the presentation rates chosen by participants when they were explicitly instructed to use either active or passive strategies to perform the task. Thus, the participant, not the experimenter, determined the presentation times for each item in the running span trial. The presentation times were recorded when participants performed the task, using either an active or a passive strategy as per specific task instructions. The data revealed a marked decrease in the presentation times during the active than the passive strategy condition. The presentation times in the active condition increased around the *n*th item establishing an updating mode for the remaining positions. This increase was absent when the passive presentation times were examined. The data in the active strategy condition resembled that found in Experiments 1 and 2, providing converging evidence for the markers of the serial-updating process.

4.2 Introduction

Experiment 2 provided the first comprehensive analysis of cognitive demands associated with different rates of presentation in running span. The investigation demonstrated that participants shift from a relatively effortless passive mode to a more demanding active mode in running span when the rate of presentation is slowed. This strategic shift was inferred using a divided attention method, in which reaction times (RTs) in a concurrent task were examined as a function of the presentation rates in running span. Results were reinforced by the recall and strategy data and were in line with previous research (Bunting et al., 2006; Hockey, 1973). Together, the data suggested that a serial-updating process taxed cognitive resources, was specifically applied after the *n*th position in a sequence and featured a specific, invariant time course in the interval between two memory items. However, these conclusions were made on the basis of specific experimental conditions. First, the use of a particular strategy was implicitly induced through a rate manipulation. Second, the index of resource demands was derived from a concurrent task rather than obtained directly from running span itself. While these findings provided an important step in the characterisation of the serial updating process, it was essential to establish that the features of serial-updating were generalisable and not the outcome of a certain methodological approach.

Morrison et al. (2016) recommended several methods for controlling and assessing participants' use of particular strategies in working memory (WM) tasks. Of these, two approaches were particularly relevant to address the need for generalisation discussed above. The first was the use of explicit strategy instructions. These can influence how participants perform a task and allow the behavioural impact of different strategies to be assessed. The second was the use of self-paced paradigms. These allow participants to control the rate of item presentation in a memory task according to their own cognitive needs rather than using a pre-determined presentation rate. These presentation times can then be used to inform an understanding of the changing cognitive load on an item by item basis throughout the trial, thereby eliminating the need of a dual-task structure that could influence the updating process.

Explicit strategy instruction has been widely employed as a method of modulating and examining cognitive strategies in working memory tasks (Carretti et al., 2007; Hockey, 1973; McNamara & Scott, 2001; Norris et al., 2019; St Clair-Thompson et al., 2010; Turley-Ames, 2003). In one study, St. Clair-Thompson et al. (2010) demonstrated that strategy instruction delivered using training games help children learn how to apply rehearsal, imagery and grouping strategies, resulting in higher performance relative to children who received no instruction. In another example, Norris et al. (2019) used strategy instruction as an experimental tool to examine cognitive operations underpinning performance in a backward recall task. Previous studies commonly suggested that backward recall is performed using a peel-off strategy, whereby an entire sequence is scanned in forward order repeatedly and the last item is "peeled off" for output with each successive retrieval (Anders & Lillyquist, 1971; Conrad, 1965). To test this, Norris et al. instructed participants to use a peel-off strategy while performing backward recall and determined that a negatively accelerating function of output times is a hallmark of the strategy. These response functions differed from those obtained from naïve participants who were not instructed to use any specific strategy, demonstrating that the use of the peel-off strategy is rare unless explicitly instructed. Hockey (1973) similarly made effective use of strategy instruction to understand the operations underlying performance in running span. He instructed participants to employ either a passive or active strategy while they performed running span and discovered that the two strategies were optimal at different rates of presentation. Recall performance was highest when the active strategy was applied at a slow rate and the passive strategy was applied at a fast rate, suggesting that the active strategy was more time-consuming that the passive one. The strategy instruction methodology thus provides a useful tool to gain experimental control over how participants approach a task and to compare performance associated with different strategies.

A second technique to gaining an insight into cognitive processing in memory tasks is by examining the time taken to process each stimulus. Most WM tasks are paced by the experimenter using a pre-determined rate such as one item per second. However, some studies have previously made use of a non-standard self-paced methodology. In these, the presentation of memory items takes place at a pace chosen by the participants, say by pressing a key to release the next item. For example, Engle et al. (1992) studied performance in a self-paced version of a short-term memory task, and showed that the data could distinguish participants with low and high task performance. The high-performers chose to receive information at slower rates than the low-performers. Based on this, the researchers argued that high-performing individuals strategically lengthened the intervals between items to allow more effective rehearsal of the recall items compared with their low-performing counterparts. Oberauer and colleagues employed this self-paced approach in a number of studies to advance an understanding of item-wise updating (Kessler & Oberauer, 2014, 2015). In these studies, the presentation times (controlled by participants) were used as an index of updating-related costs and to test competing accounts of operations involved in item-based updating. For instance, in one study, the number of items to be updated from an encoded set was varied and it was shown that the time taken to update was associated with the number of items to be updated (Kessler & Oberauer, 2015). On this basis, they suggested that updating was an item-wise process as it incurred cumulative costs for every to-be-updated item. Thus, self-paced paradigms enable researchers to develop greater understanding of the cognitive operations underlying observed behaviour. The length of the intervals between successive items can provide a correlate of the amount of strategic processing, with longer intervals suggesting greater or more time-intensive processing (Engle et al., 1992; Morrison et al., 2016).

Experiment 3

The aim of Experiment 3 was to further understand the updating process in running span, by bringing together the method of explicit strategy instruction with that of self-paced paradigms. Similarly to Hockey (1973), Experiment 3 sought to observe the impact of passive and active strategies in a running span task by instructing participants to use particular strategies. This technique was combined with the self-pacing approach used by Kessler and Oberauer (2014) in order to examine processing times associated with each running span strategy.

In Experiment 3, participants were asked to engage in either passive listening or active strategies as items were presented in a self-paced running span. Self-report data were gathered both to ascertain compliance with instructions and to provide a more detailed understanding of the range of strategies or top-down cognitive routines that participants may implement when they perform running span. Recall was tested in both active and passive strategy conditions. Based on previous research, greater recall accuracy was expected when participants used an active strategy relative to a passive strategy, with a specific benefit predicted for early recall positions (Broadway & Engle, 2010; Bunting et al., 2006; Hockey, 1973).

The variable of primary experimental interest was the presentation times of the memory items, controlled by the participants in the two strategy conditions. Based on the data from Experiment 1 and 2, it was predicted that the absence of active processing, maintenance or updating in the passive condition would leave memory vulnerable to decay over time. Participants were therefore expected to choose fast presentation times to offset the loss of information via decay. The engagement of strategic operations to maintain and update the target set in the active condition was expected to require additional time, leading participants to opt for slower presentation times. A specific updating process was predicted after position *n* in the running span sequence (Morris & Jones, 1990; Postle, 2003; Postle, Berger, Goldstein, Curtis, & D'Esposito, 2001; see also, Experiment 1 and 2 reported in previous chapters). Thus, in addition to a task-general difference in the presentation times between strategy conditions, it was also expected that the presentation times would be longer, particularly after the *n*th position, in the active but not passive condition.

Pilot investigations

Two pilot studies as reported in Appendix D1 were conducted to explore the utility of a self-paced task in which participants could regulate the presentation of memory items. The first pilot study examined these presentation times in a self-paced running span and compared them with those derived from a self-paced simple span. In this pilot study, no specific strategy instruction was provided. The data showed in running span, participants appeared to favour speeded presentation for non-update items and slow presentation for update items, and this trend was similar across participants. The function in simple span was less stable, exhibiting high variability across participants. Simple span thus did not appear suitable as a comparison task for running span, motivating a second pilot study. In this, a self-paced running span was administered, and participants were instructed to engage either actively or passively with the task. The presentation times found in the first pilot study for running span were replicated in the second pilot study in the active strategy condition, suggesting stable behavioural patterns. Further, the presentation times associated with the passive condition were low and position-invariant, suggesting that this strategy did not impose high processing demands. In addition to reliable data, the comparison of active versus passive strategies *within* running span offered an advantage in terms of task design. Both conditions used the same running span task structure, with the same variability in list length and target set size. In light of these advantages, the main experiment was designed on the basis of the second pilot study.

4.3 Methods

4.3.1 Participants

Twenty-two native English speakers (14 females, 8 males, 18 to 35 years old) were recruited for the study. The eligibility criteria and ethical guidelines described in Chapter 2 were also applied for this study.

4.3.2 Procedure

A within-subject design was used to investigate the effect of two strategies in running span. Each participant attended two sessions in which they completed running span using either active or passive strategies (counterbalanced order across sessions). A digit span task and a matrix reasoning task were also administered on separate sessions counterbalanced across participants. Each session typically lasted less than one hour and concluded with the completion of a strategy report.

4.3.3 Task

The running span task used a self-paced presentation procedure in which the participant pressed a key to trigger immediate onset of the next memory item. Participants completed the task twice, following different strategy instructions as given below with a counterbalanced order across participants. In each session, a total of 48 memory lists were presented. The task was divided into four blocks, and each block contained 12 trials with two presentations of list lengths 4, 5, 8, 9, 10, and 12, in random order. Inter-item intervals or presentation times were recorded. All other aspects of the running span procedure and task structure were the same as those employed in Experiment 1 (Section 2.3.3).

The instructions for the two strategy conditions were as follows:

ACTIVE: *In this task, you will hear strings of letters through the headset. These lists can stop at any point. Your task is to actively concentrate on the letters as you hear them, trying to continuously keep the last four letters in mind. You can do this by trying to rehearse the letters in your mind and use any other strategies that you might like. It is important to keep track of the last four letters at any given point as you cannot predict when the string will stop.*

PASSIVE: *In this task, you will hear strings of letters through the headset. These lists can stop at any point. Your task is to passively listen to the letters. Try not to pay any special attention to the letters or make any special effort to remember them. Avoid rehearsing them in your mind or grouping them together. It is important to listen passively to the string of letters.*

At the end of each block, a reminder of the instructions was provided. The level of instruction compliance was then recorded by asking participants to rate the extent to which they were following the instructions on a five-point scale from 1 = never to 5 = always. For the active condition, mean compliance = 3.9 (SD = .9) and for the passive condition, mean compliance = 3.5 (SD = 1.0).

4.3.4 Strategy reports

At the end of each session, participants completed a strategy questionnaire reporting the frequency with which they employed one or more of eight listed strategies. The strategies were: (a) passively receiving the letters, (b) rehearsing the letters as they were presented, (c) keeping up with the last four letters as they were presented, (d) grouping the letters by separating them into sets of particular sizes, (e) group the letters according to their meaning (e.g. abbreviation), (f) forming a mental image of the letters, (g) elaborating the letters by forming words, sentences or stories, and (h) generating a rhythm by tapping hands or feet. Participants rated the frequency with which they used each strategy on a four-point scale, ranging from almost never (=0) to almost always (=3).

An additional open question regarding strategies allowed participants to describe in their own words how they performed the task in both sessions. The open strategy reports (Appendix D2) were coded into strategy types by two independent raters. A broad set of strategy categories was employed as listed in Table 4.4. These included the strategies tested in the frequency questionnaire as well as other strategies described by participants after pilot testing. If a participant reported a strategy that was not previously listed by the experimenter, a new category was created. Inter-rater agreement was 87%, and if there was disagreement in the strategy coding between raters, the reports were re-examined together to reach a consensus.

4.3.5 Statistical power

An a-priori power analysis was conducted using G-Power (Faul et al., 2007). For this, an effect size was computed based on the difference between means in inter-item intervals measured during the two strategy conditions in the pilot participants (Appendix D1). With Cohen's d = 1.53, alpha = .05 and a priori power = .95, it was determined that a sample size of eight would be sufficient to run this experiment. A sample size of twenty participants was considered more appropriate, albeit slightly conservative, as it might be that the pilot data overestimated true effect size (Thabane et al., 2010).

4.3.6 Analysis plan

Presentation time data were screened for responses faster than 200 ms following the onset of an item (Van Zandt & Townsend, 2014). Further, within each condition, presentation times that deviated from individual means by more than 2.5 standard deviations and individuals who deviated from the mean presentation time by more than three standard deviations were removed. The remaining presentation time data were analysed using a 2x2 ANOVA to examine the effect of two factors: condition (active vs passive) and position (early positions up to position four vs late positions from position five onward in the sequence). In the case of an interaction, planned comparisons assessed if the difference between strategies was greater for late positions than early positions.

Recall data were analysed using a 2x4 ANOVA to examine the effect of two factors: condition (active vs passive) and recall position (one to four). In the case of an interaction, planned comparisons assessed if the difference in recall accuracy between strategies was greater for early recall positions than late recall positions. Finally, errors in recall were also examined but statistical analyses were not conducted due to relatively low frequency (mean error < 10%).

The mean frequency ratings of the eight strategies were compared between the two conditions in separate non-parametric Wilcoxon signed ranks tests.

4.4 Results

Table 4.1 provides the mean age, digit span, and matrix reasoning score of the sample as well as a summary of the performance in the running span task across both strategy conditions.

Table 4.1

Participant characteristics, presentation times and recall accuracy in both strategy conditions in Experiment 3.

		Mean	SD
Gender		14 f, 8 m	
Age		24.91	5.01
Recall accuracy[1]	Active strategy	.87	.11
	Passive strategy	.73	.17
Presentation time (s)[2]	Active strategy	2.29	1.11
	Passive strategy	.76	.28

[1] Recall scored as a proportion of items retrieved in the correct serial position

[2] Presentation times were the inter-item intervals chosen by participants

4.4.1 Presentation time

Three participants were not included in the analysis as their presentation times deviated more than three SDs from the group mean during the passive strategy condition.

A 2x2 ANOVA was conducted to examine whether there was a difference between strategy conditions across early positions (one to four) and late positions (five onwards) in the memory list. There was a main effect of strategy, $F(1,18) = 35.36$, $p < .001$, $\eta_p^2 = .66$, as well as position, $F(1,18) = 30.18$, $p < .001$, $\eta_p^2 = .63$. There was also a significant interaction between condition and position, $F(1,18) = 30.4$, $p < .001$, $\eta_p^2 = .63$.

Planned comparisons showed that the difference in presentation times between the strategy conditions was larger at late positions, $t(18) = 5.87$, $p < .001$, mean difference =

2.47 s, than early positions, $t(18) = 5.47$, $p < .001$, mean difference = .69 s. While this was in line with predictions, an inspection of the data suggested that the relative increase in presentation times in the active condition emerges at position four rather than position five as anticipated (Figure 4.1). Post-hoc pairwise comparisons (Table 4.2) showed this was indeed the case. The difference between conditions increased from position one to four, and it was then maintained for the subsequent list positions.

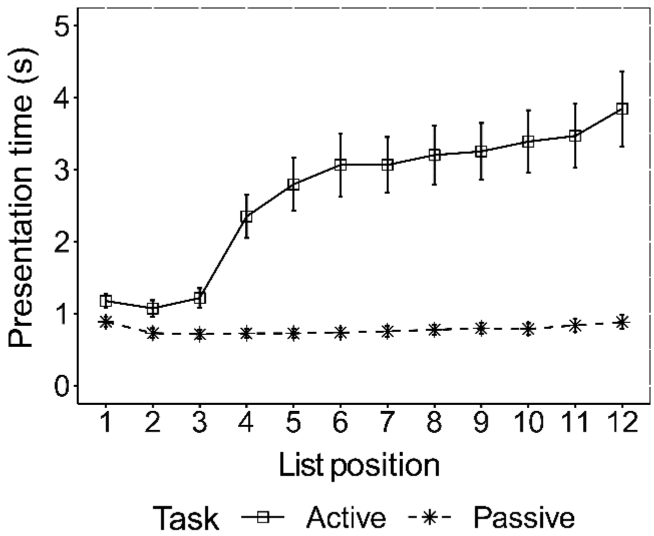

Figure 4.1 Mean presentation times at each position in the running span list in Experiment 3. Presented separately for active strategy (solid line) and passive strategy (dashed line) conditions. Error bars represent the standard error of the mean.

Table 4.2

Post-hoc pairwise comparisons of presentation times between strategy conditions for each serial position in running span in Experiment 3.

	Presentation times in active versus passive strategy				
Position	*Mean difference (s)*	*t*	*df*	*p*	*Cohen's d*
1	.29	3.48	18	***.003***	.84
2	.34	4.35	18	**<*.001***	1.18
3	.50	4.77	18	**<*.001***	1.39
4	1.62	5.49	18	**<*.001***	1.59
5	2.07	5.61	18	**<*.001***	1.60
6	2.33	5.25	18	**<*.001***	1.55
7	2.31	5.91	18	**<*.001***	1.66
8	2.43	5.82	18	**<*.001***	1.64
9	2.46	6.10	18	**<*.001***	1.70
10	2.60	5.82	18	**<*.001***	1.68
11	2.63	5.82	18	**<*.001***	1.61
12	2.96	5.43	18	**<*.001***	1.69

Note: Bold text denotes significant effects at the $p < .05$ level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method.

4.4.2 Recall

Recall data are summarised in Figure 4.2. No outliers were detected.

4.4.2.1 Accurate recall

Memory performance was compared between the active and passive conditions across the four recall positions in a 2x4 ANOVA. There was a main effect of condition, $F(1,18)$ = 53.89, p < .001, η_p^2 = .75, and a main effect of position, F(1.8,32.35) = 55.91, p < .001, η_p^2 = .76. A significant interaction between strategy and position was also found, F(1.99,35.74) = 29.33, p < .001, η_p^2 = .62. Planned comparisons showed that, relative to the passive condition, the benefit of an active mode was largest at the first recall position and decreased with each subsequent position as predicted (Table 4.3).

4.4.2.2 Recall errors

As in previous experiments, recall errors were divided into transpositions (i.e. items presented in the list but recalled at incorrect positions), intrusions (i.e. items recalled but not presented in the list), and omissions (i.e. positions at which participants omitted to recall any item by saying ‘blank’ or ‘space’). The errors committed in the task were too low to warrant statistical analyses, so here a brief discussion is provided. As presented in Figure 4.2, transpositions were the largest source of errors in the active strategy condition, while in the passive strategy condition participants committed both transpositions and omission errors. This was particularly the case for the first two target positions.

Transpositions across positions in the running span sequence were also examined (Figure 4.3). At the first two recall positions, transpositions were symmetrically located around the peak in the active but not the passive condition. The latter exhibited a backward migration of the error, such that an item presented in the position preceding the target position was more likely to be incorrectly recalled than that following the target.

Table 4.3

Pairwise comparisons of recall accuracy[1] across strategy conditions at each position in Experiment 3

	Recall accuracy in active versus passive strategy				
Position	*Mean diff*	*t*	*df*	*p*	*Cohen's d*
1	.27	8.04	18	***< .001***	2.42
2	.16	5.79	18	***< .001***	1.63
3	.10	4.59	18	***< .001***	1.22
4	.03	2.39	18	**.03**	.62

Note: Recall was scored in terms of the proportion of items recalled in the correct serial positions. Bold text denotes significant effects at the $p < .05$ level, bold italicised text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method.

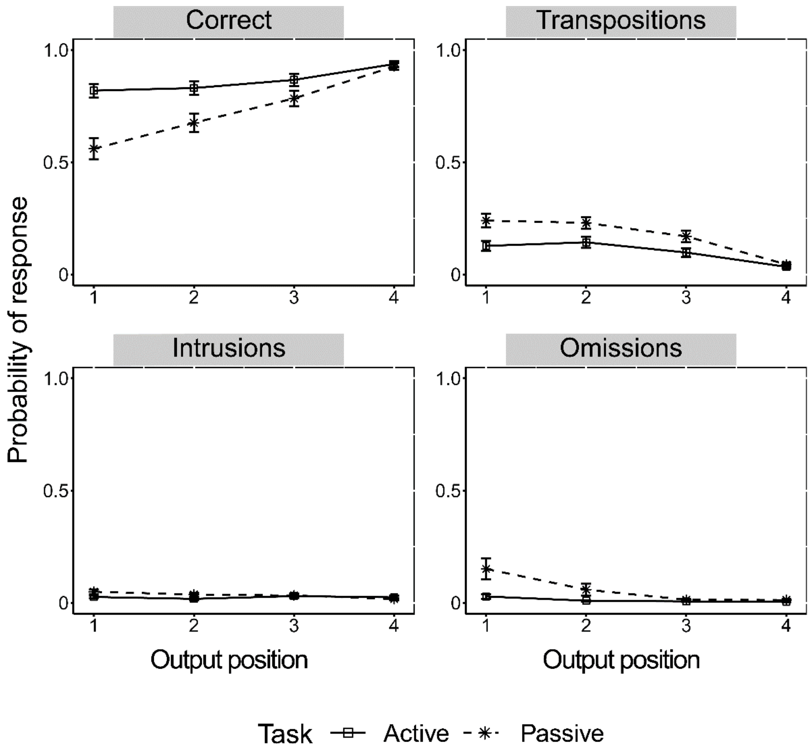

Figure 4.2 Recall data in Experiment 3. Presenting the mean likelihood of (a) correct recall, (b) transpositions, (c) item intrusions, and (d) recall omissions) at each recall position in the running span list, presented separately for active strategy (solid line) and passive strategy (dashed line) conditions. Error bars represent the standard error of the mean.

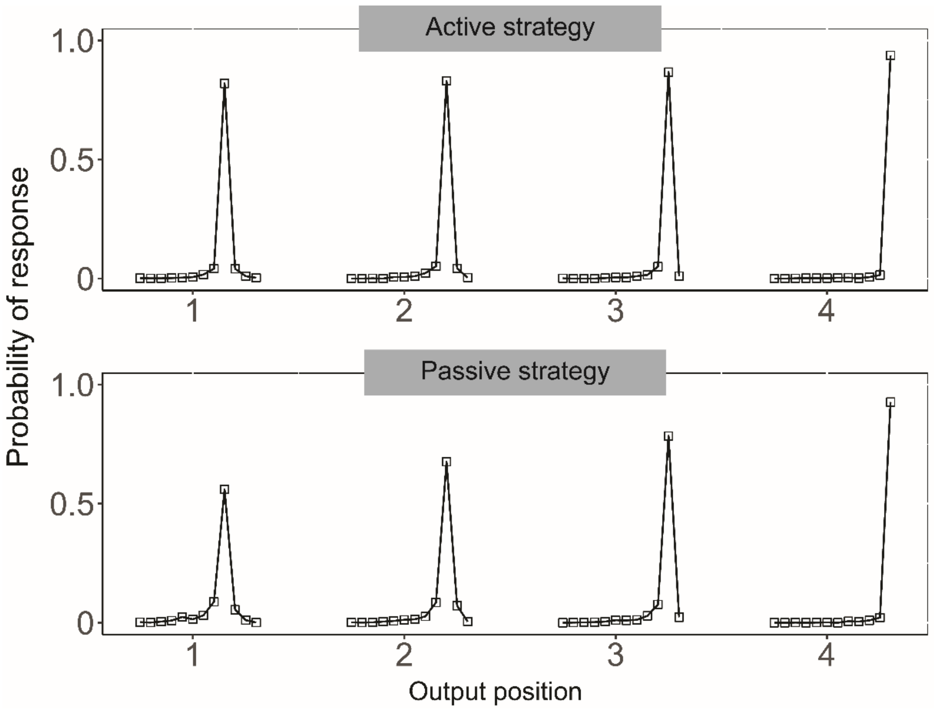

Figure 4.3 Transposition gradients for both strategies in Experiment 3. At each output position on the horizontal axis, the peaks represent the correct recall of items, flanked by the likelihood of incorrect recall of items from neighbouring positions. Error bars represent the standard error of the mean.

4.4.3 Strategy

The self-reported frequency data for the use of the eight possible strategies in running span are illustrated in Figure 4.4. Participants reported employing fewer strategies in the passive strategy condition (mean frequency rating = 0.73) compared with the active condition (mean frequency rating = 1.47). The strategies associated with active processing of presented items, such as rehearsal, chunking, elaboration, etc. were more frequently employed in the active condition than the passive condition. Passive listening was the only exception to this pattern as it was most frequently reported in the passive strategy condition. Non-parametric Wilcoxon signed ranks tests comparing the self-reported frequency ratings between the two conditions showed that they were significantly different for all strategies, all *p*s < .05, except rhythm. In addition to these, open-ended strategy reports obtained from participants were also coded as summarised in Table 4.4. No statistical analyses were carried out on these data.

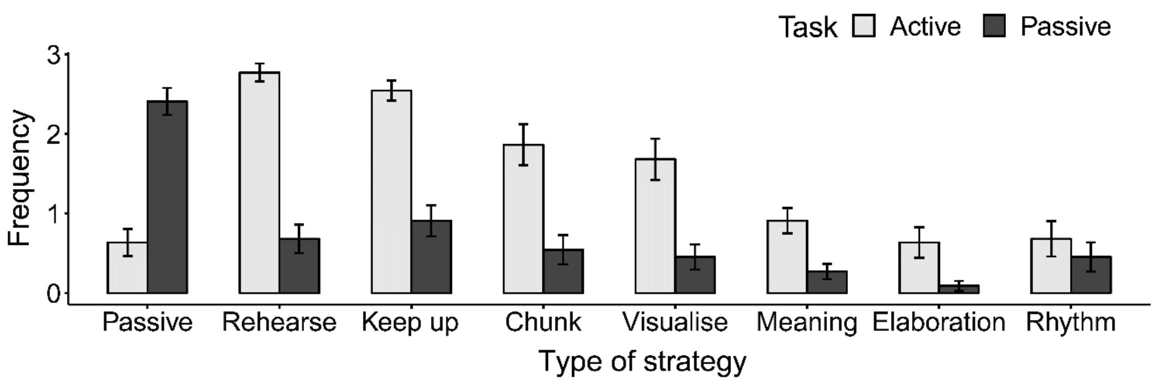

Figure 4.4 The self-reported frequency of use for eight strategies in the two running span conditions. Error bars represent the standard error of the mean.

Table 4.4

Coded self-report strategy data in both strategy conditions in Experiment 3. Participants using different types of strategies when asked to use an active or passive mode to perform the running span task. These were obtained from qualitative reports and are not included in the statistical analysis presented above.

Strategy category	Active (%)	Passive (%)
Rehearsal	59	9
Updating	45	0
Grouping	23	14
Mental imagery	18	4
Slow presentation	14	0
Semantic elaboration	14	9
Visual aid, e.g. keyboard	9	4
Embodied/kinaesthetic *	9	4
Concentration	9	0
Counting	9	0
Intonation/stress *	9	0
Tapping/rhythm	4	4
Phonetics	0	4
Speeded presentation	9	32
Single-item listening or rehearsal	4	22
Distraction e.g. look around room	0	45
Suppression of active strategies	0	36

Note: The categories marked with an asterisk (*) were added to the list of strategies after the reports were examined.

4.5 Discussion

In Experiment 3, participants were explicitly instructed to perform running span using either a passive or an active strategy, while regulating the presentation times of individual memory items. In line with predictions, participants chose longer presentation times in the active than passive condition. This finding indicates that additional operations are required in an active strategy, making it more time-consuming than a passive strategy. Within the active strategy condition, it was found that participants required less time at early than late positions in the sequence. This reinforces previous claims that participants engage in an updating process once the presented sequence becomes longer than the number of items to-be-remembered (Morris & Jones, 1990; Postle, 2003; Postle et al., 2001). The presentation time functions in the two conditions diverged markedly at position four, and those in the active condition established a new heightened but stable baseline for the remainder of the sequence. There was no position-specific increase found in the passive condition. This time course closely resembled that found in the slow-paced running span tasks in Experiment 1 and Experiment 2 and adds further weight to the proposal that slower presentation rates allow participants to apply active strategies as these are more cognitively taxing (Bunting et al., 2006; Hockey, 1973).

It is important to note that the divergence of the presentation time functions at position four was contrary to expectation, as it was one position before the expected start of updating in the list. While Experiments 1 and 2 showed a trend of increased resource demands at the fourth position, they significantly diverged from the baseline only at the fifth position. It may be that the use of a different methodology in this experiment influenced the timing of the process and caused it to shift forward by one position. As such, one possibility is that this finding may reflect pre-emptive updating, whereby participants attempted to discard the earliest item and selectively hold only the last three items in memory. In other words, a target sequence *ABCD* could be updated as *BCDx* to allow the next (fifth) item to be encoded efficiently. Such an early update could be disadvantageous if the sequence was terminated immediately after the fourth position. However, since there was an 83% likelihood in the task of receiving at least one more item in the sequence, such pre-emptive updating could be a beneficial strategic choice.

A second possibility is that the increase in presentation times at position four may reflect a preparatory process. It may be that the target set was organised in a particular format that enabled the upcoming episodes of updating. For instance, it may be that the sequence *ABCD* is segmented such that the oldest item is held separately from the remaining three items in an *A-BCD* format. This could facilitate the clear identification of the to-be-discarded item in the next updating cycle and thus support an efficient cognitive routine for serial-updating. In support of this hypothesis, some participants reported using stress, intonation or finger-mappings to segment the recall set into a one-and-three item format (e.g. ID 585, Table D2 in Appendix D2). As participants in this experiment could self-pace the items, they perhaps took the time to impose such a structure *before* the first updating episode was triggered at position five.

The recall data showed higher accuracy in the active than passive condition, indicating that participants successfully capitalised on the strategic license of the active condition. The difference between strategy conditions appeared to vary over the four serial positions, with larger differences observed for the first than the last two positions. This effect was primarily driven by the changing recall performance across positions in the passive strategy condition, while the performance in the active condition appeared relatively stable. In other words, active strategy imparted a similar benefit across all four positions, while the passive strategy was particularly beneficial for later positions. These findings were in line with the account suggesting passive listening relies on a sensory trace of recent items (Bunting et al., 2006; Cowan et al., 2005; Hockey, 1973). The impaired recall of the earlier items in the passive condition could reflect a decay in their representations, whereas the relatively flat recall function in the active condition suggest the application of cognitive operations to counteract the decay.

The relative difference in presentation times in the active versus passive condition could provide an insight into the nature of these protective operations. The additional time taken by participants in the active than passive condition beyond the fourth position was approximately 2.4 seconds. If updating occurs by unbinding them from current positions and rebinding them to new positions as proposed by the item removal account, then the present data indicate that the time needed to update is approximately 600 ms *per item*. This is consistent with previous estimates of the time course of updating, albeit in different paradigms (Ecker, Lewandowsky & Oberauer, 2014; Lewis-Peacock et al., 2018; Oberauer, 2001, 2018). Oberauer and colleagues found that, in retro-cue and complex span tasks, it took up to two seconds to remove three items, suggesting that the removal time per item was approximately 667 ms. Thus, it could be that memory updating in the running span task is accomplished using a series of item-wise updates, each taking less than a second to complete.

It is unclear whether the alternative computational models of running span considered in Section 1.3.2 could account for these data as they lack the temporal specification. One of these models is the *suppression account* that suggests that the to-be-discarded item in a target set is suppressed during the updating, while the remaining items are rehearsed to maintain their activation levels in WM. A second *overwriting account*, based on the *n*-back model by Chatham et al. (2011), assumes that serial order is represented circularly, such that the start- and endpoints of the encoded set are the same. This model advances a process by which the oldest item is removed and replaced by the newest item, thereby overwriting it. The data obtained in this experiment suggest that the process of suppression and overwriting in each respective account would take approximately 2.4 seconds. In the absence of detailed information about the temporal requirements of either process, the models can neither be supported nor opposed. However, future theoretical development of the models could incorporate the temporal constraints provided in this study.

These computational solutions could also be empirically tested using a self-paced running span task with varying *n*. The position in favour of item-wise updating would predict an increase in mean presentation times as the target set increases because each item will need to be updated one at a time. In contrast, both the suppression and the overwriting accounts would predict a single updating operation irrespective of the number of items to be recalled. The item-based updating account will be supported if the presentation times increase as the *n* in running span increases. Assuming a linear relationship, the anticipated increase in mean presentation times would be approximately 600 ms for every increase in the value of *n*. If the presentation times are insensitive to variation in *n*, it would lend support to the alternative single operation accounts. The present experiment did not allow for these predictions to be tested as a fixed value of *n* was employed, but these hypotheses could be tested in future studies.

Insights into the process of serial-updating were also gained by examining how participants engaged with the task demands and how they tried to keep track of the *n* relevant items. Self-report data from the present study revealed considerable variation in how participants approach the task. In the active condition, a large proportion of participants reported rehearsing items and updating the target set by trying to eliminate the irrelevant item from memory and focusing on the relevant items. Some participants also reported the use of grouping (e.g. using formats such as A-BCD) and mental imagery to support their recall. In the passive condition, participants primarily reported having to suppress these active strategies by distracting their attention away from items and instead attending to objects in the room or their internal cognitive or body experiences (e.g. mind wandering or rubbing their knees). Notably, some participants reported strategically choosing slow presentation rates in the active condition to improve performance, particularly from the fifth item onwards (e.g. ID 579, Table D2 in Appendix D2). They also reported releasing the items as fast as possible in the passive condition so that they could avoid memorising (e.g. ID 585, Table D3 in Appendix D2) or using mental imagery (e.g. ID 579, Table D3 in Appendix D2). Although previous studies have examined strategy use in running span (Hockey, 1973; Morrison et al., 2016), this is the first study to gather qualitative accounts of strategies underpinning both the active and the passive processing modes.

Together, the data in Experiments 1 to 3 converge to demonstrate that running span can be performed using both passive and active strategies. Results using both divided attention and self-paced paradigms consistently showed that updating is time- and resource-consuming, and that it is characterised by a specific time course. The serial-updating process applied only during active strategies appears to be prompted by the arrival of the first item to-be-updated, although some evidence for a pre-updating preparatory process was also found. These particular features of the data are not simply consequences of a dual task mode and are also evident in a single task structure. Given the slow presentation times preferred in the active condition, it is unsurprising that fast presentation rates fixed by the experimenter induce passive listening as these provide a small fraction of the time that they would ideally need for updating. However, an outstanding question remains as to whether these temporal functions are specific to running span or may be generalised to other serial updating tasks. Experiment 4, presented in the following chapter, addressed this question by exploring the resource demands during *n*-back tasks.

Chapter 5. Cognitive demands in *n*-back

5.1 Overview

The results from the first three experiments presented in this thesis consistently showed that serial updating in running span costs resources and time. Experiment 4, presented in this chapter, extended the investigation to another serial updating paradigm in order to test the generalisability of the demand features observed with running span. The *n*-back task was chosen as the candidate serial updating task. This was based on previous research suggesting a common updating component between running span and *n*-back (Bäckman et al., 2017; Chatham et al., 2011; Dahlin et al., 2008). Other experimental studies and cognitive analyses have also highlighted that the similarity between the two tasks might be limited due to the differences in the memory test, i.e. serial recall in running span versus item recognition test in *n*-back (Gathercole, Dunning, Holmes, & Norris, 2019; Zhao et al., 2018). The aim of Experiment 4 was thus to test if the resource demands of serial updating in *n*-back are similar to those involved in running span, or whether the demand profiles are task-specific even if the tasks rely on similar updating processes.

The time course of demands was examined in two *n*-back conditions, employing the same divided attention approach used in previous experiments. The *n*-back conditions, *1*-back and *4*-back, had distinct memory updating but similar item recognition demands. The time courses suggest that the process of serial updating follows the item recognition judgement. Also, a comparison of the demand functions observed in *4*-back and running span (using data from Experiment 1) indicated the involvement of a similar serial updating process in both tasks with comparable demand profiles. This demonstrated that, despite varying procedures of testing memory performance, the resource demands of serial updating were generalisable across at least two paradigms.

5.2 Introduction

The experiments reported so far have examined the demands in running span, a complex working memory (WM) task that requires updating of encoded information and its serial order. Another task imposing similar updating requirements is *n*-back. In this, participants are presented with lengthy sequences of memory items and their task is to judge whether the more recent item matches the one presented *n* positions back in the sequence. The *n*-back task is widely used in neuroimaging to identify neural correlates of WM in general and WM updating more specifically (Awh et al., 1996; Hartley et al., 2001; Owen et al., 2005).

N-back and running span are considered 'cognate' procedures as they share similar cognitive demands (Ruiz & Elosúa, 2013). In running span, *n* items have to be recalled in serial order once the list is over. Similarly, in *n*-back, the latest *n* items are relevant in the memory task, albeit not for a complete serial recall. Instead, the first of this set of *n* items (the *n*-back item) has to be compared with the most recent item. It is important to note that even though only the *n*-back item is immediately required for these recognition judgements, the remaining items in the target set are required in turn for recognition judgements later in the sequence. As both tasks require the maintenance of the last *n* items at any given point in the course of the trial, it is plausible that they might recruit similar operations to keep track of the latest set of target items.

Consistent with this view, Chatham et al. (2011) developed a computational model of *n*-back and reported that the same computational principles and architecture could also support running span performance. This model assumed a periodic serial representation in which only *n* serial positions were maintained. The first *n* items were encoded at *n* positions respectively. Thereafter, the process of encoding cycled back to the beginning of the list such that the latest item would overwrite the earliest item at each updating step. Section 1.3.2 provides a full description of the model. Applied to running span, the model resulted in a similar updating solution, bolstering the proposal of a common updating process shared by the tasks.

Further evidence for the overlap between running span and *n*-back was provided by a neuroimaging study of cognitive training (Dahlin et al., 2008). In this study, a variety of adaptive running span tasks, i.e. tasks that become increasingly more difficult, were administered across multiple sessions and the benefit of this training on *n*-back measures was examined. It was observed that trained participants improved their *n*-back performance compared with participants who did not undergo any training. Using functional magnetic resonance imaging (fMRI), the study also found that both tasks were associated with similar fMRI activity in the left striatum. On this basis, the authors proposed that the process of updating, common to both running span and *n*-back, is supported by striatal resources and cognitive training enhances this striatal signal and transfers the benefit to the *n*-back task. This proposal was reinforced by positron emission tomography (PET) findings reported by Bäckman et al. (2017). In the PET study, both running span and *n*-back tasks were associated with striatal dopamine release and the dopaminergic activity was heightened during *n*-back performance following running span training. Together, the two studies suggested that the involvement of an overlapping component in running span and *n*-back tasks and the authors claimed that this is the process of serial updating.

However, the PET study described above also tested intensive, adaptive cognitive training using running span tasks but found no evidence for the transfer of benefits to *n*-back performance (Bäckman et al., 2017). It is important to note that both the studies by Dahlin et al. (2008) and Bäckman et al. (2017) described above lacked an active control group. The role of nonspecific factors impacting on performance, such as changes in general cognitive engagement, motivation or social contact, cannot be accounted without an appropriate control group (Simons et al., 2016), making it premature to draw strong conclusions on the basis of their results. Further, a more recent study administered adaptive *n*-back training to participants and found that there was no improvement in running span performance as a result of training-related gains in the *n*-back task (Zhao et al., 2018). These mixed results from cognitive training studies highlight the fact that the cognitive demands of the two tasks are not identical, despite the similarity in their neurochemical profiles.

So what might be the critical difference between the two paradigms? Cognitive analysis suggests that an important source of difference is the way in which memory is tested. Running span simply requires the serial recall of *n* items at the end of the list. In contrast, the *n*-back task tests the recognition of a single item. Participants have to compare the current item in the sequence (probe item) with the *n*th item back from the sequence end (target item) and respond differently to matches and mismatches. Each stimulus presentation in *n*-back, therefore, is a prompt not only to encode and update the target set as in running span, but also to perform a recognition judgement (Rac-Lubashevsky & Kessler, 2016).

It could be that the two tasks rely on a common process of serial updating, which is embedded in dissimilar cognitive routines (Gathercole et al., 2019). According to the framework proposed by Gathercole et al., cognitive routines are novel, temporary and sometimes idiosyncratic solutions generated by coordinating cognitive processes in a particular sequence. These enable us to perform complex and unfamiliar tasks that cannot be performed using our already existing cognitive apparatus. The authors argued that running span and *n*-back have distinct higher order task structures (i.e. preparing for upcoming serial recall at the end of the list versus continuously engaging in a memory recognition test for each new item). They speculated that these necessitate different cognitive routines and that the process of memory updating, even if similar, would be applied in the context of different processes in running span and *n*-back

It is important to note that the two components of *n*-back, namely serial updating and item recognition, are not easily disentangled. A possible solution advanced by Rac-

Lubashevsky and Kessler (2016) involved the use of reference-back, a novel task that allowed the updating and recognition judgements to be studied separately. This task uses two types of trials, namely reference and comparison trials. In the reference trials, participants update their WM (i.e. store the latest item as the new 'reference') as well as make recognition judgements (i.e. compare latest item with previous 'reference'). In the comparison trials, only recognition judgements are required. This novel paradigm is thus said to isolate an updating component by comparing performance in the reference and comparison trials. However, the updating requirements in reference-back and *n*-back are dissimilar. Reference-back involves a *single* item change such that an old item is simply replaced by a new item. This is in contrast to the complex requirement of serial updating in *n*-back, in which one item is discarded while the rest are repositioned. Therefore, reference-back updating is more akin to item updating and it is thus an inappropriate vehicle to gain an insight into the serial updating process.

Following Watter et al. (2001), an alternative method of separating the updating and item recognition components in *n*-back could be to test multiple *n*-back conditions. Watter et al. used electroencephalography (EEG) and studied the P300 event related potential while participants completed *n*-back tasks. Watter et al hypothesised that the P300 component would mark the item comparison and decision process, and it should thus be found across multiple *n*-back tasks that vary in the level of n. In this with this prediction, Watter et al. showed that the latency of the P300 component was similar across *n*-back tasks, suggesting a common time course of recognition. It was further found that the amplitude of the P300 component varied as a function of memory load (= n). As the value of n increased, the P300 amplitude decreased. The authors proposed that this reflected a recruitement of resources away from the memory decision, thus a decrease in amplitude, which were applied to maintain and update WM. Critically, they showed that the updating process, but not item recognition, was sensitive to the value of *n*. This pattern could be exploited if the contribution of the two components is to be distinguished.

To summarise, there is conflicting evidence from task analyses, computational modelling, training studies and theories of training-related transfer regarding the degree of overlap between running span and *n*-back (Bäckman et al., 2017; Chatham et al., 2011; Dahlin et al., 2008; Gathercole et al., 2019; Ruiz & Elosúa, 2013; Zhao et al., 2018). On one hand, similar computational architecture and striatal neurochemistry appear to underpin performance in both running span and *n*-back tasks. On the other hand, contrasting theoretical perspectives and mixed evidence of transfer cast doubt on the proposed overlap. These instead suggest that the distinction in the memory test between the two tasks leads to different cognitive solutions that cannot be applied across tasks. It also appears that isolating the updating and recognition judgements within *n*-back is non-trivial but comparing profiles across multiple *n*-back conditions could present a way forward.

Experiment 4

The aim of Experiment 4 was to establish whether the *n*-back and running span tasks involved the same serial updating process. One way to shed light on the level of similarity between cognitive tasks is to compare their general cognitive demands. Craik et al. (1996) used a divided attention approach to show that the encoding and retrieval processes in long-term memory imposed different demands. Adopting the same approach, Experiment 1 and 2 showed that resource demands in a slow-paced running span were different from those of serial recall in simple span (Experiment 1) and passive listening in a fast-paced running span (Experiment 2), thereby identifying demand features specific to the updating process. In the same way, the divided attention approach could be used to examine the resource demands of *n*-back, by applying a concurrent choice reaction time (CRT) task. The concurrent CRTs supply a metric of the resource demands of *n*-back and these could be compared with the CRT-running span data obtained in Experiment 1.

To briefly recap, the results from Experiment 1 and 2 demonstrated that the process of serial updating in running span was composed of two elements. One was a general mode of updating triggered in the trial specifically when the number of encoded items exceeded the required number of target items. A second, localised process, supporting the updating of the target set with each new item, occurred with a latency of 1000 ms from item onset. Based on these results, a few alternative outcomes could be anticipated in the present experiment even though it was primarily exploratory in nature. First, it could be that identical demand profiles occur in *n*-back and running span. If this were the case, *n*-back demands would exhibit a general rise following the presentation of the $n+1$th item in the trial and an additional, specific increase around 1000 ms from item onset. This outcome would suggest that there is a close correspondence between the tasks not only in the involvement of serial updating but also in its timing and evolution through the entire course of the trial. A second possibility was that serial updating demands in *n*-back resemble running span but differ in their timing. It may be that the presence of the item recognition decision in *n*-back impacts the timing of the updating process, which occurs with a time lag compared with that in running span. In this case, the latency of the item-level demand would be greater than 1000 ms. If instead the latency is shorter than 1000 ms, it could suggest that updating occurs prior to the anticipated item recognition decision. A final consideration, of course, was that the demand profiles in *n*-back and running span are entirely distinct. This outcome would suggest a complete absence of overlap between the tasks and suggest that the process of updating is a task-specific solution that lacks generalisability.

It was crucial to identify the demands specific to the serial updating process in *n*-back and distinguish them from the recognition decision. Following Watter et al. (Watter et al., 2001), the comparison of two *n*-back conditions could enable a distinction between the two elements for the task. In this experiment, therefore, the resource demands in *1*-back and *4*-back were measured, by combining each *n*-back task with the CRT task. While both conditions were anticipated to involve recognition judgements, the updating load was higher in *4*-back than *1*-back (Watter et al., 2001). Based on this, it was assumed that similarities in the CRT functions would reflect the demands of the item comparison and decision process, while differences would indicate the demands due to memory updating. It was also expected that the recognition accuracy would be higher in the *1*-back than the *4*-back task based on previous investigations of the impact of *n* on memory performance (Jaeggi et al., 2010; Juvina & Taatgen, 2007; Kane et al., 2007; Mackworth, 1959; Oberauer, 2005; Szmalec et al., 2011).

In summary, Experiment 4 examined the time course of resource demands in *n*-back and tested the generalisability of the serial updating process across tasks. As in Experiments 1 and 2, the divided attention approach was used in which the CRT task was concurrently applied with two *n*-back conditions. The resultant CRTs supplied a metric of the resource demands of *n*-back. These *n*-back demand profiles were compared to those observed with running span in Experiment 1 and it was expected that similarities, if any, between the *n*-back and running span tasks could reflect the common serial updating process.

Pilot studies

Pilot studies tested whether the divided attention approach employed in Experiment 1 and 2 and the self-paced procedure used in Experiment 3 could be administered to measure resource demands of the *n*-back task. Appendix E1 provides details of these pilot studies. The first pilot study demonstrated that a self-paced *n*-back task in which participants could determine the stimulus presentation rates was not a suitable procedure for this study. In this task, a single keypress triggered the onset of the next stimulus as well as indicated a same/different judgement comparing the latest item that was presented in the *n*-back position. Thus, the presentation times did not allow a distinction between the time taken for memory updating and recognition judgements. In the second pilot study, the divided attention approach was used to measure resource demands associated with two *n*-back conditions, *1*-back and *4*-back. Most task parameters were based on the running span task administered in Experiment 1 to maximize procedural comparability between *n*-back and running span. The pilot data showed a greater cost of dividing attention during *4*-back than *1*-back in line with general predictions based on increased memory load. Also, the time course of the more demanding *4*-back condition increased around position *n* (=4) as might be expected based on previous running span investigations. Unexpectedly, the demands reduced towards the later positions in the *n*-back sequence instead of maintaining a steady, high level as observed in running span. While such a profile could be specific to *n*-back, it was also possible that it demonstrated that the task parameters optimised for running span were unsuitable for *n*-back. To eliminate the latter possibility, the task parameters were modified and aligned with typical *n*-back investigations for the main experiment described below.

5.3 Method

5.3.1 Participants

Thirty participants (9 male, 21 female, mean age = 25 years, SD = 4.6 years) participated in the study. The eligibility criteria and ethical guidelines described in Chapter 2 were also applied for this study.

5.3.2 Procedure

The study used a 2x2 design investigating the effect of two factors, *n*-back condition and attentional load. *N*-back condition had two levels (*1*-back versus *4*-back) and attentional load also had two levels (single and dual load). Both factors were manipulated within-subject, i.e. all participants completed both single and dual task versions of each *n*-back condition.

Each participant provided informed consent to participate in the study and completed two sessions with the different *n*-back conditions on separate days. Three tasks were administered in each session, a choice reaction time (CRT) task, the respective *n*-back task, and a dual load condition in which both the CRT task and *n*-back were performed simultaneously. Each session started with a practice block that was then followed by five experimental blocks. The order of the tasks was fixed within a block (CRT, *n*-back, followed by the dual task). A strategy questionnaire was then administered before the session was concluded. Each session typically lasted for an hour.

5.3.3 Tasks

5.3.3.1 *N*-back

Participants were presented long sequences of items and were required to detect matches between the current item and that presented *n* positions ago in the sequence. Two *n*-back tasks were administered with $n = 1$ and $n = 4$ in separate sessions. Both tasks used the same procedure for sequence generation and stimulus presentation. Each *n*-back task consisted of five blocks with four sequences each. Sequences consisted of $20 + n$ letters, out of which there were six targets and $14 + n$ non-targets. The following eight letters comprised the stimulus set, *B, F, H, K, M, Q, R,* and *X*. Each letter was presented for a duration of 800ms and followed by a silent interval for 1600 ms. Spoken presentation was used and the stimuli used in the present study were the same as those used in Experiment 1 (Section 2.3.3.1). In this experiment, participants were instructed to say “match” aloud as soon as they detected a target match. Keypresses, as typically used in *n*-back tasks, were not possible in the dual load condition due to the responses (via keypresses) required in the concurrent choice reaction time task (see below). Thus, spoken responses were employed in both the single and dual load conditions on the *n*-back task. The use of spoken responses also maintained parity with the other experiments in this thesis.

In the *1*-back condition, participants were required to respond every time the most recent stimulus was the same as the item presented one position back in the list. In the *4*-back condition, participants were required to similarly respond every time the latest stimulus matched the item presented four positions back in the list. No responses were required in the case of no match. The spoken responses were manually time-stamped by the experimenter to assess recognition accuracy. If participants responded ‘match’ accurately to one item before the onset of the next item, they were regarded as a ‘hit’; else they were recorded as false alarms. Typically, *n*-back responses are also analysed in terms of latencies. This was however not undertaken in the present investigation as the timing of the response data suffered from poor precision due to the combined use of spoken responses and manual timestamping.

5.3.3.2 Choice reaction time task

Choice reaction times (CRTs) were measured using the same visual four alternative choice task as in Experiments 1 and 2. See Section 2.3.4 for detailed description of the task.

5.3.3.3 Dual load condition

In the dual load condition, the *n*-back and CRT tasks were simultaneously applied (Figure 5.1). The protocol to implement the dual load condition in this experiment was the same as that used in Experiment 1, with the exception of using *n*-back instead of running span (Section 2.3.5 for further details).

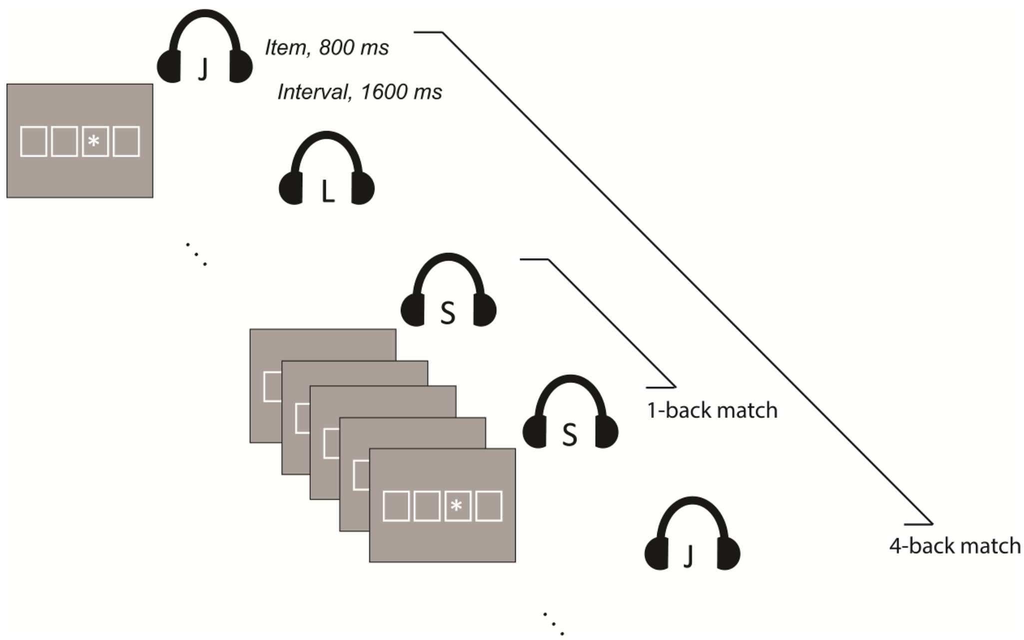

Figure 5.1 An illustration of the dual-load condition with a simultaneous application of (auditory) *n*-back and (visual) CRT task in Experiment 4. Memory items presented every 2400ms; CRT task was participant paced, so the number of CRT stimuli presented between each memory item varied contingent on participant reaction times.

5.3.4 Strategy reports

At the end of each session, participants completed a strategy questionnaire reporting the frequency with which they employed one or more of eight listed strategies. The strategies were: (a) passively receiving the letters, (b) rehearsing the letters as they were presented, (c) keeping up with the last one/four letters as they were presented, (d) grouping the letters by separating them into sets of particular sizes, (e) group the letters according to their meaning (e.g. abbreviation), (f) forming a mental image of the letters, (g) elaborating the letters by forming words, sentences or stories, and (h) generating a rhythm by tapping hands or feet. Participants rated the frequency with which they used each strategy on a four-point scale, ranging from almost never (= 0) to almost always (= 3).

5.3.5 Statistical power

The kind of temporal analysis used in this study has not been applied to *n*-back before. The effects obtained in previous running span experiments were therefore relied on. In Experiment 1 and 2, the observed *post hoc* power of the reported effects was greater than .90. Thus the same sample size of 30 as in these experiments was also used here.

An additional *a-priori* power analysis was conducted using G-Power (Faul, Erdfelder, Lang, & Buchner, 2007) using the effect size obtained from pilot data. This was computed based on the difference between single and dual RTs measured in the second pilot in the *4*-back condition (Appendix E1). This determined that a total sample size of 13 would be sufficient to run this experiment based on an estimate of the effect size from Cohen's d = 1.11 (as determined within G-Power), alpha = .05 and a-priori power = .95.

5.3.6 Analysis plan

Data were trimmed and screened for outliers as in Experiment 1.

Reaction time. Experiment 4 explored the temporal characteristics of resource demands during *1*-back and *4*-back conditions, indexed by the CRTs from the concurrent task. The analyses were guided by those in previous experiments but were not identical due to differences in the number of updating items per *n*-back condition. At the task level, a 2x2 ANOVA was used to examine the differences in concurrent RTs as a function of two factors: attentional load (single versus dual CRTs) and *n*-back condition (*1*-back versus *4*-back). At the trial level, a 2x2 ANOVA was used to examine the difference in CRTs associated with update and non-update positions in both *1*-back and *4*-back conditions. To maintain comparability of the number of data-points between the *n*-back conditions, this analysis compared the CRTs during the last non-update position with the first update position in each sequence. In other words, the analysis compared item 1 (non-update) versus item 2 (update) in the *1*-back condition and item 4 (non-update) versus item 5 (update) from the *4*-back condition. This however did not consider the data from the remaining non-update and update positions in the *4*-back condition. To track how the CRTs changed over the course of the trial, paired sample t-tests compared CRTs between each pair of adjacent positions in the *4*-back sequence. At the item level, the interval between successive item onsets (2400 ms) was segmented into six 400 ms bins. Again, due to the differences in number of data-points between the *n*-back conditions, non-update and update positions were analysed separately. Accordingly, two separate 6x2 ANOVAs were used to examine CRTs as a function of bin and *n*-back condition for each position category. In case of significant interactions, post-hoc analyses were used to find the source of the interaction.

To compare the resource demands during *n*-back and running span, the data from the *4*-back condition were compared with those from Experiment 1. For both sets of data, the interval between updating items was segmented into six bins of 400 ms and a 6x2 ANOVA was conducted. CRTs were examined as a function of two factors: bin (within-subject) and task (between-subject).

Recognition. A 2x2 ANOVA compared the recognition accuracy as a function of two factors: attentional load (single versus dual load) and *n*-back condition (*1*-back versus *4*-back). Similarly, a 2x2 ANOVA examined the false alarm rate across loads and *n*-back conditions. A significant main of *n*-back was predicted and post hoc tests were used to explore any interaction effects.

Strategy use. Non-parametric Wilcoxon signed ranks tests compared the mean frequency ratings between the *1*-back and *4*-back condition, separately for each strategy.

5.4 Results

Table 5.1 summarises findings for the CRT and *n*-back tasks as a function of the *n*-back condition (*1*-back and *4*-back) and attention load (single and dual load).

Table 5.1

Participant characteristics and performance in concurrent CRT task and *n*-back task, for each load and *n*-back condition in Experiment 4.

		1-back	*4*-back
RT (ms)	Single	417 ± 48	414 ± 44
	Dual	428 ± 52	517 ± 96
Recognition accuracy [1]	Single	.88 ± .07	.44 ± .22
	Dual	.86 ± .08	.39 ± .22
False alarm rate [2]	Single	.02 ± .02	.10 ± .06
	Dual	.02 ± .02	.11 ± .07

[1] *N*-back accuracy was scored as the percentage of targets correctly identified.

[2] False alarms were scored as the percentage of lures incorrectly identified as targets.

5.4.1 Reaction times within *n*-back

Outlier screening of RTs resulted in the exclusion of one participant from the following analyses. The CRT data under dual load condition are summarised in Figure 5.2 (across trials) and Figure 5.3 (across the item-interval).

5.4.1.1 Task level

A 2x2 ANOVA compared single and dual CRTs across the *1*-back and *4*-back conditions. There was a significant main effect of attentional load, $F(1,28) = 48.08$, $p < .001$, $\eta_p^2 = .63$, a significant main effect of *n*-back condition, $F(1,28) = 18.39$, $p < .001$, $\eta_p^2 = .40$, and a significant interaction between the two factors, $F(1,28) = 29.91$, $p < .001$, $\eta_p^2 = .52$. Paired sample t-tests, summarized in Table 5.2, showed that dual CRTs were significantly slower during *4*-back than *1*-back, mean difference = 88 ms. RTs for *1*-back and *4*-back conditions did not differ significantly under the single load condition.

Table 5.2

Paired sample t-tests of CRTs between each *n*-back condition, computed for single and dual loads separately in Experiment 4.

	1-back versus *4*-back			
	Mean diff (ms)	*t (28)*	*p*	*Cohen's d*
Single CRTs	2	.332	.742	.06
Dual CRTs	88	5.22	***<.001***	1.05

Note: The data analysed here include all CRTs after trimming and outlier correction, across all lists within each task (for dual condition). Bold text denotes significant effects at the $p < .05$ level, bold italicized text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method.

5.4.1.2 Trial level

A 2x2 ANOVA compared CRTs during the presentation of the first update item (position 2 in *1*-back and position 5 in *4*-back) with the preceding non-update item (position 1 in *1*-back and position 4 in *4*-back) in both *n*-back tasks. There was a significant main effect of item, $F(1,28) = 4.95$, $p = .03$, $\eta_p^2 = .15$, reflecting longer RTs at the update than non-update item (mean difference = 13 ms). There was also a significant main effect of *n*-back, $F(1,28) = 32.97$, $p < .001$, $\eta_p^2 = .54$, reflecting longer RTs during the *4*-back than the *1*-back (mean difference = 126 ms). There was no significant interaction between position and *n*-back, $F(1,28) = 1.3$, $p = .26$, $\eta_p^2 = .04$.

Paired sample t-tests compared RTs between each successive pair of positions to track the change in CRTs within the *4*-back trials. These are detailed in Appendix E2. In brief, the results showed that the CRTs increased from the first to the fourth position, with no significant change from position four onwards.

5.4.1.3 Item level

A 6x2 ANOVA compared CRTs, during non-update positions, across the six bins of the item interval (400ms each) between *1*-back and *4*-back task. There was a significant main effect of bin, $F(1.9,54.7) = 5.66$, $p = .006$, $\eta_p^2 = .17$, and a significant main effect of *n*-back condition, $F(1,28) = 28.17$, $p < .001$, $\eta_p^2 = .50$. There was no significant interaction between task and bin, $F(2.7,75.2) = 2.6$, $p = .06$, $\eta_p^2 = .09$.

A corresponding 6x2 analysis compared CRTs during the updating positions as a function of the six bins between the *n*-back conditions. There was a significant main effect of bin, $F(2.2,60.9) = 19.20$, $p < .001$, $\eta_p^2 = .41$, and a significant main effect of *n*-back condition, $F(1,28) = 28.49$, $p < .001$, $\eta_p^2 = .50$. There was also a significant interaction between task and bin, $F(2.1,58.9) = 6.86$, $p = .002$, $\eta_p^2 = .20$.

Post hoc analyses, summarised in Table 5.3, showed that at update positions in *1*-back, there was an increase in RT from the first to the second bin, no change between the second and the third bin, and a subsequent decrease from the third bin onwards. A similar profile was observed in *4*-back, with an additional, significant increase from the second to the third bin. Also, the magnitude of these differences in CRTs between successive bins were larger in *4*-back than in *1*-back.

5.4.2 Reaction times in *n*-back versus running span

Figure 5.4 summarises the CRTs at the item level in both *4*-back and running span. For this data from Experiment 4 (*4*-back) is presented alongside the data from Experiment 1 (running span).

A 6x2 ANOVA compared CRTs during update positions across the six bins of the interval between items (within subject factor) between *4*-back and running span tasks (between-subject factor). A significant main effect of bin was found, $F(2.6,147.0) = 14.5$, $p < .001$, $\eta_p^2 = .21$, such that there was an overall increase in RT from the first to the third bin and a decrease from the third bin onwards. There was no significant effect of task, $F(1,56) = .19$, $p = .67$, $\eta_p^2 = .003$, and there was no significant interaction between bin and task, $F(2.6,147.0) = .99$, $p = .42$, $\eta_p^2 = .02$.

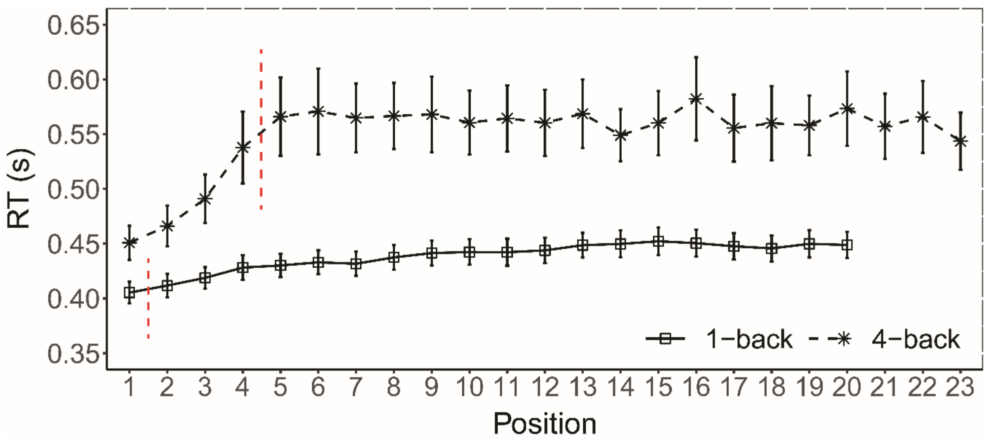

Figure 5.2 Mean concurrent CRTs across list positions for each n-back task in Experiment 4. Note that there were fewer positions in 1-back than *4*-back due to shorter list lengths. RTs associated with the final position across lists are not displayed here, see text for data exclusion. The vertical dotted red lines indicate the separation of update and non-update positions, as per the n in each task. Error bars represent standard error of the mean.

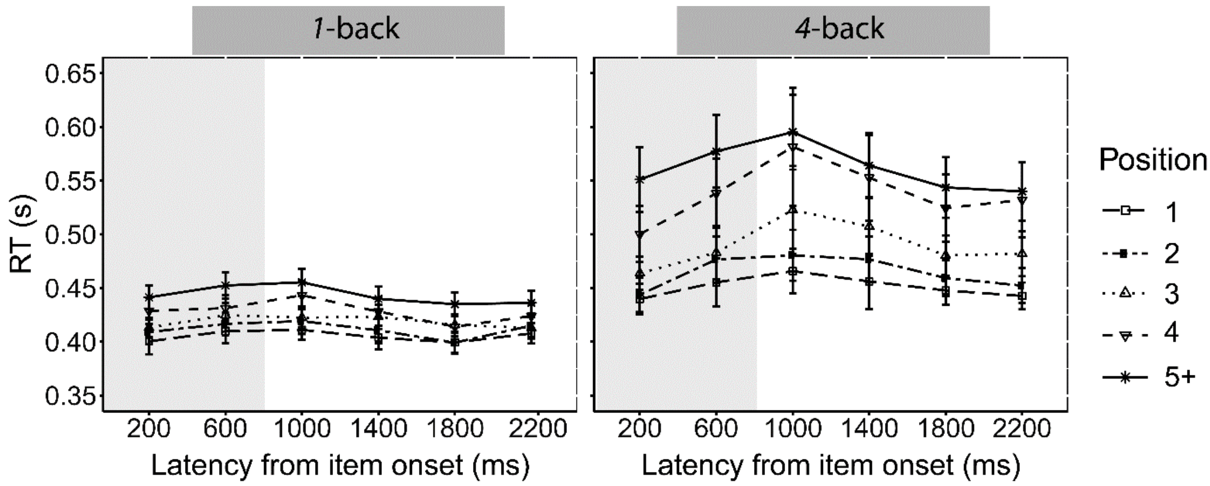

Figure 5.3 Mean concurrent CRTs as a function of latency from item onset of memory item across positions, plotted separately for 1-back (left) and *4*-back (right) in Experiment 4. Please note that while the data are illustrated per position, the analysis collapsed items into non-update and update positions (see text for more). The first 800 ms represent the duration of the item presentation (shaded in grey), followed by a 1600 ms silent inter-item interval (unshaded). Error bars represent standard error of the mean.

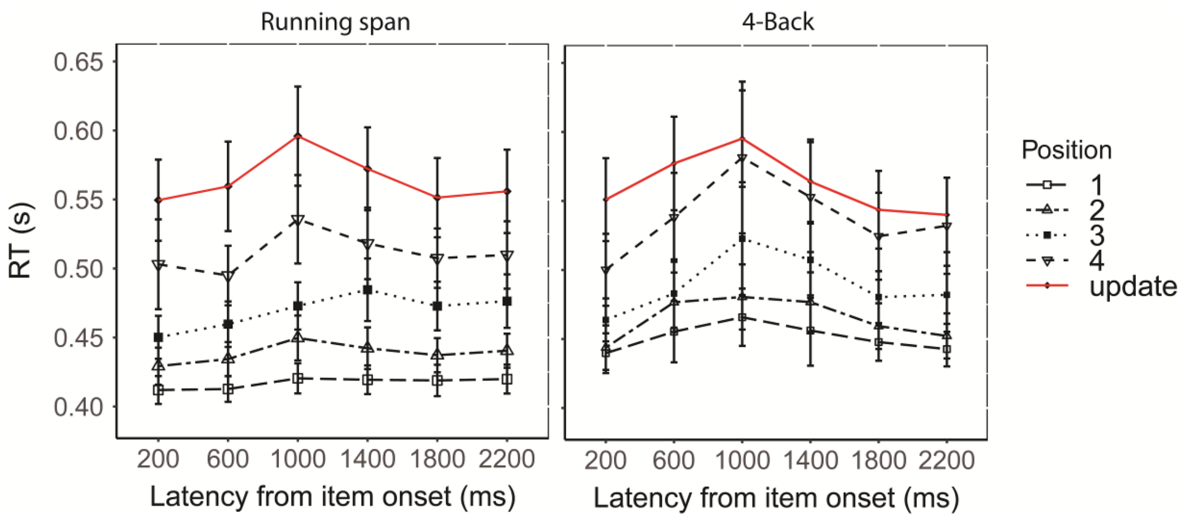

Figure 5.4 Mean concurrent CRTs as a function of latency from item onset of memory item across positions, plotted separately for running span (left; data from Experiment 1) and 4-back (right; data from Experiment 4). Please note that while the data are illustrated per position, the analysis was only conducted for update positions (highlighted as the red line; see text for more). The first 800 ms represent the duration of the item presentation (shaded in grey), followed by a 1600 ms silent inter-item interval (unshaded). Error bars represent standard error of the mean.

Table 5.3

2x2 ANOVAs to examine the interaction between bin and task for each pair of consecutive bins between *1*-back and *4*-back during the update positions, and pairwise mean differences in RT where applicable in Experiment 4.

	Interaction effect between bin and task				Pairwise mean diff in RT	
	F	*df*	*p*	η_p^2	*1*-back	*4*-back
Bin 1 vs 2	10.74	1,28	***.003***	.28	***9***	***26***
Bin 2 vs 3	6.61	1,28	**.02**	.19	3	***23***
Bin 3 vs 4	7.91	1,28	***.009***	.22	***- 13***	***- 30***
Bin 4 vs 5	8.8	1,28	***.006***	.24	***- 5***	***- 23***
Bin 5 vs 6	.45	1,28	.51	.02	.	.

Note: The data analyzed here include dual CRTs during update positions after trimming and outlier correction extracted from *1*-back and *4*-back. Pairwise analyses were conducted only if the interaction between bin and position for that bin pair was significant. An increase in RT between consecutive bins is indicated in positive mean difference values, while a decrease is indicated in negative values. Bold text denote significant effects at the $p < .05$ level, bold italicized text indicate significance effects after adjusting for multiple comparisons using the Bonferroni method.

5.4.3 Recognition performance

Recognition accuracy and false alarm rates are summarised in Table 5.1.

A 2x2 ANOVA compared recognition accuracy as a function of attention load (single vs dual) and *n*-back condition (*1*-back vs *4*-back). There was a significant main effect of load, $F(1,29) = 18.62$, $p < .001$, $\eta_p^2 = .39$, in that accuracy was higher for single than dual load conditions, mean difference = 4%. There was also a significant main effect of *n*-back condition, $F(1,29) = 166.49$, $p < .001$, $\eta_p^2 = .85$, whereby accuracy was higher for *1*-back than *4*-back condition, mean difference = 46%. There was no significant interaction between the two factors, $F(1,29) = 3.89$, $p = .06$, $\eta_p^2 = .12$.

A similar 2x2 ANOVA compared false alarm rates as a function of attention load (single vs dual) and *n*-back condition (*1*-back vs *4*-back). There was a significant main effect of *n*-back, $F(1,29) = 47.63$, $p < .001$, $\eta_p^2 = .62$; there were fewer false alarms in *1*-back than *4*-back task, mean difference = 9%. There was no significant main effect of load, $F(1,29) = 1.51$, $p = .23$, $\eta_p^2 = .05$, and no significant interaction between the two factors, $F(1,29) = 1.12$, $p = .29$, $\eta_p^2 = .04$.

5.4.4 Strategy

Participants reported employing fewer strategies in the *1*-back condition (mean frequency rating = 0.69) compared with the *4*-back condition (mean frequency rating = 1.12). With the exception of passive listening, all other strategies were most frequently employed in *4*-back than *1*-back (Figure 5.5).

Non-parametric Wilcoxon signed ranks tests showed significantly greater use of strategies such as keeping track of the last four items, elaboration, and chunking, in *4*-back and the greater use of passive listening in 1-back, *p*s < .05. In contrast, rehearsal, visualisation, grouping by meaning, and rhythm were used to the same extent in both tasks.

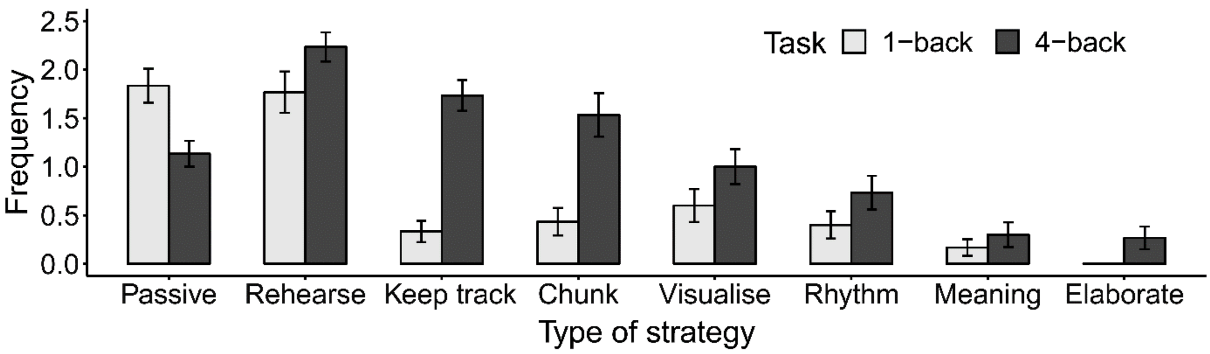

Figure 5.5 Mean frequency ratings of self-reported strategy use for each n-back condition in Experiment 4. Participants rated their use of each strategy from 0 (= almost never) to 3 (= almost always), see text for full statements.

5.5 Discussion

The objective of Experiment 4 was to identify the resource demands of serial updating in *n*-back and contrast these with running span demands. Using the divided attention approach, two *n*-back tasks were administered concurrently with the CRT task and the concurrent CRTs supplied an index of resource demands. To distinguish the demands associated with *n*-back updating and recognition judgements, the concurrent CRTs associated with *4*-back and *1*-back were compared. In line with predictions, the CRTs were considerably slower in the former. This pattern indicated greater levels of cognitive demands when participants had to maintain and update four items rather than a single one, consistent with the EEG data reported by Watter et al. (Watter et al., 2001). There was also a marked disparity in memory performances across the two tasks. Performance was significantly poorer in *4*-back than *1*-back in terms of both recognition accuracy and false alarms, in line with previous studies demonstrating a decrease in performance with an increase in the value of *n* (Jaeggi et al., 2010; Juvina & Taatgen, 2007; Kane et al., 2007; Mackworth, 1959; Oberauer, 2005; Szmalec et al., 2011).

Across both *n*-back tasks, the CRTs associated with the first update item were greater than the final non-update item, but the magnitude of this effect of updating did not vary between tasks. This could suggest that the activation of an updating mode when the first to-be-updated item is encountered demands similar resources irrespective of the *n*-back condition. It may also be that there is indeed a difference between the resources required to trigger an updating mode in *4*-back versus *1*-back, but it was not apparent in the data. This is because the CRT increase in *4*-back was observed at position four, rather than position five (= n+1) at which they were anticipated.

This early elevation in processing demands was also found in Experiment 3 that examined the (self-paced) presentation times in running span. In both experiments, the variable of interest (presentation times in Experiment 3 and concurrent CRTs in Experiment 4) successively increased until position four (= *n)* rather than position five (= *n+1).* Two possible explanations considered in the context of the previous experiment could also be applied to the present data. One possibility is that participants preemptively update their target set at position four. In other words, they might discard the first encoded item from a sequence, e.g. *ABCD*) and continue holding the last three items in the format *BCDx* to allow for the fifth item to be encoded efficiently. While this may be a useful strategy in running span in which the likelihood of a recall test following position four was quite low, it is not as effective in *n*-back. This is because the memory of the first item is *always* tested in *n*-

back, due to the recognition judgement required at position five to compare it to the first item. The second, more likely possibility is that the increase in CRTs at position four reflects a preparatory process in which participants anticipate the first recognition judgement at position five and attempt to ready their WM for a quick decision process. For instance, it may be that segmenting the sequence *ABCD* in the format *A-BCD* allows the relevant *n*-back to be rapidly identified and easily compared with the most recent item.

In addition to the rise in CRTs at position four, the CRTs were also elevated 600 ms from the onset of update items in both *4*-back and 1-back. This was the interval in which item presentation was still ongoing. Following this, the CRTs in *4*-back, but not *1*-back, continued to increase, with a maximum level observed at a latency of approximately 1000 ms. The CRTs subsequently decreased over the remaining interval returning to the updating baseline before the onset of the next item in both *1*-back and *4*-back. The increase at 600 ms could reflect the recognition judgement as it was common to both *n*-back tasks. The specific elevation at 1000 ms specifically found in *4*-back could reflect the serial updating process, as it was more demanding in *4*-back than 1-back. Together, these data suggest that the recognition judgement in *n*-back occurs while stimulus is being presented, prior to memory updating.

The CRT increase at around 1000 ms in the *4*-back task may also reflect resources required to resolve the conflict between a familiarity and recollection signal (Oberauer, 2005; Szmalec et al., 2011). It has been previously argued that *1*-back task does not require conflict resolution as the most familiar item is likely to be the last item encountered and thereby enabled rapid access to the target *1*-back item (Oberauer, 2005). In fact, it may be that in *1*-back every new item simply replaces the last, and only, item held in WM without recruiting any update process. As such, it would resemble a relatively quick and effortless memory substitution or reset process, as in modified span in Experiment 1 (see also, Kessler & Meiran, 2008). This interpretation was corroborated by the strategy data, whereby active demanding strategies were more common in *4*-back and passive listening in *1*-back.

A comparison of the CRT functions between running span and *n*-back revealed no difference. This outcome favours the overlap account that the two tasks rely on the same process of updating (Bäckman et al., 2017; Chatham et al., 2011; Dahlin et al., 2008a; Ruiz & Elosúa, 2013b). However, these data are inconsistent with the finding that *n*-back involves an item recognition process evident 600 ms from stimulus onset. It is surprising that the recognition judgement (required only in *n*-back and not running span) was not apparent when CRT data were compared between the two tasks. It could be that the two components were indistinct, perhaps even overlapping in time within *4*-back, such that the statistical analyses could only detect them in comparison with *1*-back but not running span.

These data warrant further investigation and highlight the need for a different means to separate the *n*-back task components. Instead of comparing two *n*-back conditions, it may also be more useful to decompose *n*-back into its constituent components (item comparison and memory updating) to develop a greater understanding of the task demands. Such an approach was used by Rac-Lubashevsky and Kessler (2016) who developed a novel reference-back task. As argued in the introduction of this chapter, the reference-back task however fails to capture the demands of serial updating as it only requires a single item replacement rather than systematic changes to the serial positions of already-encoded items (p. 131). A modification of the task could present a way forward. Instead of a single item target set, the reference-back could involve a target set of *n* items. In this case, each cycle of updating would involve the discarding of the first target item and repositioning of the remaining ones, as in typical serial updating paradigms. Future investigations could compare the demands of updating trials with those in which WM is unchanged and thus isolate the demand characteristics specific to *n*-back updating.

In summary, Experiment 4 found a close resemblance in the resource demands of *n*-back and running span, supporting the hypothesis that the two tasks rely on the same process of serial updating. The data additionally suggested that the memory recognition judgement in *n*-back precedes the memory update process. Suggestions for further investigations were made to clarify the timing and unique demands of the two *n*-back components, while also allowing more specific tests of the updating process in both *n*-back and running span

Chapter 6. General discussion

The overarching aim of this thesis was to provide the first empirical study of how serial order information is updated in working memory (WM). Our current understanding of working memory updating is derived from studies of item-based updating, in which serial order representation is unchanged. In contrast, the current conceptualisations of serial updating to date have been largely limited to descriptive accounts. Here, a series of systematic and well-controlled experiments examined the serial updating process in two paradigms, using two methodological approaches.

Experiment 1 applied a divided attention approach to examine the resource demands of serial updating. The demands of running span, a serial updating task, were compared with those of non-updating tasks such as standard serial recall, with the aim of identifying temporal features in the data specific to the updating process in running span. Using the same divided attention approach, Experiment 2 measured the resource demands in three running span tasks that varied in their presentation rates. The objective of this study was to provide conditions that either facilitated (slower rates) or impeded (faster rates) the process of serial updating and observe the impact of these conditions on resource demands. Experiment 3 used a self-paced running span paradigm in which participants controlled the presentation times for the memory items. Participants were explicitly instructed to use either active updating or passive listening to perform the task. The goal of this study was to use converging operations to test whether presentation times varied as a result of strategy. Experiment 4 assessed the extent to which the results of the investigation in running span could generalise to another serial updating task. For this, the *n*-back task was used, which differs from running span in its memory test procedure. While running span requires immediate serial recall of the latest *n* items at the end of a sequence, *n*-back tests item recognition in which participants have to compare the most recent item with the *n*-back item and make same/different judgements. The divided attention approach was again employed to record the resource demands of memory updating and item recognition in two *n*-back variants, *1*-back and *4*-back, and compare these with running span data from Experiment 1.

The key results and interpretations of the four experiments are summarised below, followed by a discussion of the theoretical implications of the work. Limitations of the studies and avenues for future research are then identified. Finally, the thesis concludes with a summary of the key outcomes of this research.

6.1 Summary of results

The data in Experiment 1 demonstrated that running span placed higher demands on cognitive resources than standard serial recall, reflected by the slower choice reaction times (CRTs) in a concurrent task. Concurrent CRTs in running span increased when the encoded sequence became longer than the target number of items, with the greatest impact observed approximately 1000 ms from stimulus onset. These results indicate that an updating mode that imposes heightened demands is activated following the *n+1*th item in running span. An additional process triggered by each new update item was found, which could reflect the application of cognitive mechanisms such as repositioning, removal or suppression. These data represent the first empirical investigation of the resource demands incurred during running span to date, providing a temporal precision lacking in our current understanding of the serial updating process based on recall data (Morris & Jones, 1990; Postle, 2003; Postle et al., 2001).

Experiment 1 also found that the concurrent CRTs were faster in modified span, a task that required participants to periodically discard the *entire* contents of WM and restart encoding a new list, compared with serial updating in running span. Interestingly, CRTs *decreased* following the presentation of update items in modified span, in contrast to the increase observed in running span. The serial position functions in running span and modified span also differed. Recall accuracy in running span increased monotonically with each output position, whereas recall in modified span showed both the recency and primacy components of a standard serial position function. These differences in CRT and recall data suggest that it is faster and easier to discard all previously encoded items and start encoding afresh than it is to discard some items while retaining others. These observations are consistent with previous studies using item-based updating paradigms (Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014; Kessler & Meiran, 2008), demonstrating that the distinction between a complete versus partial change in WM is also found in serial updating paradigms.

Experiment 2 then established that resource demands of running span increased as the rate of presentation decreased. At presentation rates slower than one item per second, the time course of concurrent CRTs was identical to that observed in Experiment 1. At faster rates, the onset-locked CRT increase disappeared. This indicates that the process of updating is time-consuming and can only be applied during relatively lengthy intervals between successive items, as suggested by previous research (Bunting et al., 2006; Hockey, 1973). Recall performance was impaired by the dual load of the concurrent CRT task *only* in the slow-paced running span task. This suggests that the running span and concurrent CRT task competed for common, capacity-limited executive resources at the slow but not medium or fast rates, likely due to the specific application of updating in the former. This is in line with Botto et al. (2014) who showed a reduction in performance when available resources were distributed between running span and a prospective memory task. The self-report strategy data reinforced this interpretation, with participants using strategies that required actively rehearsing, grouping and updating items more frequently during slower item presentation. In contrast, at faster rates, there was a shift towards a relatively effortless passive listening approach.

Experiment 3 shifted from a dual-task paradigm with fixed presentation rates to a self-paced presentation procedure, in which they were instructed to use explicit updating strategies. Participants paced items at a slow rate of ~3.5 seconds per item when instructed to use an active updating strategy and shifted a more rapid presentation of ~1 second per item while using the passive listening strategy. The overall difference in presentation times between the two strategy conditions increased at the fourth (= *n*th) position in the sequence and maintained a stable level for the remaining positions. A relatively flat serial position function was found for recall in the active strategy condition. For the passive strategy condition, a monotonic serial position function was found that corresponded to with those typically found in running span tasks (Bunting et al., 2006; Morris & Jones, 1990; Palladino & Jarrold, 2008; Ruiz et al., 2005). The disparity in recall accuracy usually found between early and late output positions in running span thus reduced when participants were allowed as much time as needed to complete the task. The disparity re-emerged when participants were advised to use passive listening. The strategy data derived from open self-reports as well as frequency ratings showed substantial variability in the use of active strategies in running span, reinforcing findings from Experiment 2. These converging results suggest that the demands of the serial updating process are independent of methodological aspects of the investigation, such as dual- versus single-task structures and experimenter- versus participant-determined presentation rates.

Experiment 4 examined the time course of resource demands of another serial updating paradigm, the *n*-back task. The *4*-back was associated with slower concurrent CRTs and poorer recognition performance than *1*-back. This was in line with previous suggestions that an increase in the value of *n*, and thereby an increase in the working memory load, results in a decrease in performance accuracy (Jaeggi et al., 2010; Oberauer, 2005; Szmalec et al., 2011). Both *n*-back tasks exhibited heightened CRTs approximately 600 ms from the onset of update items, with a greater magnitude observed in *4*-back than *1*-back. This could reflect the common *n*-back component, namely item recognition, required in both tasks. There was an additional increase in CRTs approximately 1000 ms from stimulus onset in *4*-back, but not *1*-back. This could reflect the process of memory updating, as the updating of four items was more demanding than that of one item. Together, these data suggest that the item recognition judgements in *n*-back occur prior to updating but require further investigation. The CRTs associated with *4*-back in Experiment 4 were indistinguishable from those associated with running span in Experiment 1, suggesting that a similar updating process is involved in both tasks. Finally, the self-report strategy data showed a greater reliance on updating (among other active strategies) in *4*-back and passive listening in *1*-back. This pattern resembled the strategy use observed in running span in Experiments 2 and 3. This bolsters the conclusion that the serial updating process characterised in these experiments was generalisable across running span and *n*-back.

6.2 Theoretical understanding of serial updating

The four experiments greatly inform our understanding of the serial updating process involved in running span and *n*-back tasks. Five aspects of the data are discussed below.

First, the data provide an explicit characterisation of an active updating process distinct from passive listening. This is contrary to previous proposals that running span can only be performed by passive listening (Broadway & Engle, 2010; Elosúa & Ruiz, 2008; Ruiz & Elosúa, 2013; Ruiz et al., 2005). Instead, results instead favour Hockey's proposal that running span can be performed using either active or passive strategies (Hockey, 1973). They also establish that the choice between the strategies is determined on the basis of temporal parameters of the task, such as the rate of presentation (Botto et al., 2014; Bunting et al., 2006; Hockey, 1973).

Second, a direct comparison of running span and *n*-back tasks demonstrates similar resource demands to the common updating process. This mirrors findings on the neural level, showing common striatal involvement and dopaminergic activity across both tasks (Bäckman et al., 2017; Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008b). The present data extend previous results, showing that the time course of demands was indistinguishable between the two tasks, at least at the temporal resolution employed here. This substantiates efforts, such as by Chatham et al. (2011), to develop computational models of serial updating that could simulate performance in a number of paradigms. The evidenced overlap, of course, does not preclude the fact that there are components specific to each task that need to be assimilated into computational and cognitive models. But it does suggest that the task-specific components do not result in a complete change of the time course of serial updating.

Third, there appears to be an updating mode initialised after the encoding of the first few items (notwithstanding the specific time point at which it is activated). This mode may garner the cognitive control necessary to facilitate a switch between encoding (or maintenance) of relevant items and updating of the irrelevant items (Kessler, 2017). In this way, it may reflect a state in which the gating mechanism proposed in computational models of WM is prepared but not yet applied (Frank, Loughry, & O'Reilly, 2001; Miller & Cohen, 2001). When it is closed, the gate protects WM from interfering contents of ongoing cognition and perception but allows these influences when it is opened. Thus, the increased demand at position *n* or *n+1* could reflect a state of readiness for updating.

Fourth, the observed time course implied that each updating cycle is prompted by the presentation of a new stimulus. The process that enables the updating of outdated, irrelevant information is triggered. It is time-consuming, requiring at least one second under paced conditions even when the duration of item presentation is short. The data from un-paced conditions show that optimally the process takes approx.2.5 seconds.

What might this process involve? It could be that serial updating is accomplished by a series of individual item updates. The first to-be-discarded item could be eliminated from WM by reversing the encoding process, as suggested in the item removal model by Oberauer and others (Lewis-Peacock et al., 2018; Oberauer, 2018). The remaining items would be repositioned one at a time, shifting them one back in the target set, so that the newest item is in the last target position. The timing data from the current research suggest that these removal and repositioning operations will take approximately 600 ms each, which is consistent with previous estimates of the time taken to update (Oberauer, 2018).

Other accounts suggest that the process of serial updating involves a single operation, resulting in a change for either the first or the last item of the target set. The suppression account of updating, proposed in this thesis by extending principles of the Primacy model (Page & Norris, 1998), advances a process by which the first (oldest) item is suppressed while the remaining items are rehearsed to protect them from time-based decay. The overwriting model of *n*-back developed by Chatham et al. (2011) instead proposes that the oldest item at any given point is overwritten by the newest item. Based on the timing data obtained here, the suppression and overwriting mechanisms take 2.5 seconds each, which could be used to constrain future model development.

Finally, the multiplicity observed in the strategy data is compelling and needs to be acknowledged. Participants reported using item rehearsal, keeping up with the target set, grouping, mental imagery, and tapping among other strategies in the active processing conditions. In contrast, they reported needing to suppress these spontaneous active strategies in the passive conditions, doing so by restricting themselves to single-item listening and shifting attentional focus from the stimulus to their internal or external environment. Such strategic variability was consistent with previous reports in running span (Morrison et al., 2016) as well as *n*-back (Minear et al., 2016). Importantly, these results suggest large individual differences in how complex tasks are performed that may not be captured using only a single computational model (Norris et al., 2019). It may be that there are no existing pieces of cognitive machinery readily available to perform tasks such as running span and *n*-back. Instead, participants may have to develop idiosyncratic, flexible and possibly temporary cognitive structures when faced with these tasks (Gathercole et al., 2019) . If this is the case, the demands documented in this thesis could reflect the general cognitive resource required for flexible cognitive control, rather than a single, unitary process of updating. The attempt to identify a modal model might nonetheless still be useful, as it could identify the general course of events during updating even if the underlying mechanisms might vary across individuals.

6.3 Limitations

There are some key limitations of the present research. These are identified below along with suggestions of how they might be overcome in the future.

First, although the methods used in this thesis charted the resource demands of updating in two paradigms, they were unable to identify the specific nature of the cognitive operations involved. Therefore, the mechanisms underlying the resource demands of updating remain unclear. As described above, serial updating could involve a single mechanism, such as suppression or overwriting, or a multi-step operation, such as local, item-wise changes. As such, these models offer contrasting predictions. The suppression and overwriting accounts of serial updating require a single operation irrespective of the number of items to be recalled. In contrast, the item-wise account of updating requires an operation per item, thus predicting an increase in the time required to update with an increase in the number of item-position bindings to change. One way therefore to test these contrasting predictions is to change the target set size, *n*, in running span while using the self-paced procedure as in Experiment 3. If the presentation times are insensitive to the variation in *n*, it would support the single operation accounts of serial updating, although further work would be required to ascertain whether suppression or overwriting underpins the update process. If the presentation times instead increase with an increase in *n*, it would support the item-based account of updating. In this way, the alternative models could be tested empirically to determine the cognitive mechanisms underlying updating.

Second, the present findings failed to identify the exact position at which there is a shift from an encoding-maintenance mode to an updating mode. According to Morris and Jones (1990), an executive process to support WM updating should be activated when the *n+1*th item is presented. While data in Experiments 1-2 supported this proposal, Experiments 3-4 suggested a cognitive shift earlier in the list at position *n*. This mixed evidence could reflect an artefact of the trial structure employed in the current experiments, in which it was five to nine times more likely that participants would receive the fifth (= *n+1*th) item than they would be cued to recall the first four (= *n*) items. Participants may have thus preemptively updated the target set one position earlier to allow the next item to be encoded more efficiently. One way to test the possibility of preemptive updating is to present each stimulus with an equal, fixed probability. For example, Ecker and colleagues administered an updating task in which there was a constant likelihood of 50% that the list would be terminated after each item (Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014). This ensured an equal probability of an update and a recall event, eliminating the advantage of an early update. The current investigations could thus adopt Ecker et al.'s approach to probabilistic list termination. If the resource demands at position *n* continue to be observed, it would challenge the hypothesis that these demands are induced due to differential item probabilities. Such an outcome would instead suggest a preparatory process that readies the system for the first, upcoming episode of updating in each trial. In this way, the cognitive process underlying the resource demands at item *n* could be better understood.

The investigation of *n*-back suggested two separable task components, item recognition and serial updating, with different time courses. However, these components could not be established due to the challenge of decomposing complex tasks such as *n*-back. It is imperative for future research to use alternative trial structures to allow an examination of the individual elements of the task. The reference-back task developed by Rac-Lubashevsky and Kessler (2016) could povide a solution as it requires updating and target recognition in different trials, thus effective separating the two activities. Other hybrid tasks such as the reference-back could be developed in the future, which are better able to capture serial updating and isolate it from other components of the task.

6.4 Future research directions

Building on the foundations of the research presented in this thesis, four avenues for further investigation are outlined.

6.4.1 Testing generalisability of serial updating across task conditions

The present experiments demonstrated similar resource demands of serial updating in running span and *n*-back. To maximize comparability, the same task conditions were used across both paradigms, including the size of the target set (= *n*), rate of presentation, stimulus domain and modality. Further research could determine whether the demand functions are generalisable across other task conditions. A key finding was that the activation of a demanding, updating mode occurs at or after the fourth (= *n*th) item, indicating that it is dependent on the value of *n* used in the task. Another, less likely, possibility is that the demands set in at fourth position in the sequence regardless of the value of *n.* To rule out this possibility, demand profiles could be examined in different running span tasks with varying the value of *n*. If there is indeed an association between the timing of the updating mode and the number of items to-be-recalled, a variation in *n* should result in a corresponding shift in the serial position at which the updating mode is activated.

Similarly, further research would be required to test the generalisability of the present results to other task conditions such as other verbal stimuli (e.g. digits or words), visual presentation of verbal items (e.g. letters presented on the screen) or visuospatial items (e.g. colors or spatial locations). While it is proposed here and elsewhere (Morris & Jones, 1990) that the updating process is facilitated by general cognitive resources, it could be that the time course of updating is domain-specific. Previous research suggests that memory storage and maintenance processes might be different for verbal and visuospatial memoranda (Morey, 2018; Morey et al., 2019). This of course does not necessarily imply that the representation of serial order might vary across verbal and visuospatial content (Hurlstone et al., 2014). However, if the processes of encoding, maintenance and updating rely on the same general resources, the resultant demand characteristics for the updating of visuospatial stimuli could be different from those observed for auditory, verbal stimuli.

6.4.2 Examining resource demands of item updating

This thesis focused on serial updating. Another category of WM updating is item updating, as described in Section 1.3.1, in which the item itself undergoes a change rather than its order information. For example, the memory updating task employed by Ecker et al. (2014) taxed the item updating process. In this task, participants held *n* items (e.g. letters) in *n* spatial locations on the screen (e.g. a row of square frames in the centre of the screen). Each new update item was presented in a particular location and had to replace the item previously held in that location. This task thus involves local changes in items with no amendment that impacted the remaining items. It has been proposed that such updating is performed by first removing the older item from WM and then encoding the new item in its place (Ecker, Lewandowsky, et al., 2014; Ecker, Lewandowsky, Oberauer, & Chee, 2010; Oberauer, 2018).

Researchers have also suggested that the same process of item removal is also applicable to tasks such as *n*-back (Rac-Lubashevsky & Kessler, 2016). This hypothesis could be tested by using the same divided attention approach applied here with an item updating task. A similar time course of resource demands across different updating tasks would suggest a singular process of updating invariant to specific task demands. In contrast, if the temporal analyses of the resource demands reveal different functions for item and serial updating paradigms, the results would instead support the distinction proposed in this thesis. This would demonstrate that the assumed equivalence between different types of updating is unfounded and ensure that future investigations do not treat the two categories interchangeably.

6.4.3 Increasing temporal resolution of analysis

The divided attention approach used in Experiments 1-2 (running span) and Experiment 4 (*n*-back) allowed for the measurement of resource demands over the course of an item interval using data bins of 400 ms each. With this temporal resolution, the data reflect a *single* process of updating that demands resources 1000 ms from stimulus onset. This process could however instead be composed of multiple, distinct mechanisms, for instance to support the repositioning of consecutive items, with each mechanism associated with a local maximum in the demand function. There may also be variation in the timing of the maximum demand across individuals.

There is a thus need in future research to increase the temporal precision of the demand function, for instance by increasing the number of data points using the same experimental design presented in this thesis. This could be achieved by increasing the number of participants, trials per session, and items per trial, which would in turn enable a decrease in the size of the data bins and thereby increase the resolution of the investigation.

An important limitation of a continued reliance on the divided attention approach however is that the results could be influenced by the concurrent task and thus may not be pure measures of the updating process.

An alternative approach to achieving an increase in temporal resolution is to adopt psychophysiological techniques such as pupillometry. Previous studies show that cognitive effort or demands (rather than task difficulty per se) can be feasibly tracked using task-evoked pupil dilations measures (van der Wel & van Steenbergen, 2018). Pupil dilation is associated with the number of digits to be remembered in a digit span task, thus indexing load on short term memory resources (Johnson, Miller Singley, Peckham, Johnson, & Bunge, 2014). Also, individual differences in *n*-back performance were captured by the variation in pupil dilation (Rondeel, van Steenbergen, Holland, & van Knippenberg, 2015). In a similar manner, pupil dilation could be used to chart the variation in resource demands over the course of active updating, compared with standard serial recall or passive listening.

Brain-based measures of cognitive activity using methods such as electroencephalography (EEG) and magnetoencephalography (MEG) also present alternatives to the divided attention approach and yield neuroimaging data with rich temporal resolution. In previous studies using EEG, the P300 event-related potential has been associated with *n*-back task performance (Scharinger, Soutschek, Schubert, & Gerjets, 2017; Watter et al., 2001). For instance, Watter et al. showed that the timing of the P300 was common across several *n*-back conditions with varying levels of *n*, whereas the magnitude of the peak varied across the conditions. They argued that these results could reveal different task components, with the peak latency reflecting the recognition judgement and the magnitude reflecting the load on memory resources. Future investigations could build on this work and compare the EEG signal, such as the P300 component, between update and non-update items across both serial updating and item updating tasks. This would provide a comprehensive analysis of ongoing cognitive activity during WM updating in a variety of contexts.

6.4.4 Advancing cognitive and computational models of serial updating

A final avenue for further research motivated by the current work is the development of theoretical models enhancing our understanding of how serial order of items held in WM is updated. Data collected across the four experiments do not favor one particular model over another but provide empirical insights that could guide specific theoretical accounts of serial updating. Section 1.3.2 outlined a number of alternative computational solutions to implement serial updating, either by adapting existing models of serial recall or extending models of item updating to also achieve serial updating.

Importantly, there are two outstanding issues that need to be addressed in future research. The first is to understand whether capacity limitations of WM arise due to time-based decay of memory traces or interference between memory representations. This has been a vigorous debate in the general field of WM research for several years (Barrouillet et al., 2007, 2011; Farrell et al., 2016; Lewandowsky & Oberauer, 2009, 2015; Oberauer & Lewandowsky, 2008, 2014; Portrat et al., 2008). Resolving this debate is critical to conceptualising active updating. The decay-based models assume that WM capacity is limited as each item suffers a decay over time. In this case, updating would be modelled as an active process that selectively counteracts the decay of the relevant items, say by rehearsing (Burgess & Hitch, 1992, 1999; Page & Norris, 1998) or refreshing (e.g. Barrouillet et al., 2004; Barrouillet & Camos, 2012). In contrast, if WM capacity limitations are due to increased interference or competition between representations, then updating would be considered a process that actively resolves the interference, perhaps by removing or inhibiting the outdated items (Lewis-Peacock et al., 2018; Oberauer, 2018). It could of course also be the case that both decay and interference contribute to limit the capacity of WM, which would have to be incorporated into the updating model.

A second question is whether serial updating is a single-step process or follows an item-by-item approach in a multi-step process. Single-step updating could occur in models that use relative rather than absolute representations of serial order. For instance, the Primacy model proposed by Page and Norris (1998) proposes that items are held in WM with different levels of activation and these activation levels decrease with each serial position in the list. This model could incorporate serial updating by suppressing the first target item of the encoded sequence (encoded with the highest activation level) while the new item is simply added to the end of the sequence with an appropriate (low) level of activation. On the other hand, multi-step updating would be required in models that rely on an absolute coding of order, e.g. by binding items to positions (Burgess & Hitch, 1999; Kessler & Oberauer, 2014, 2015; Oberauer et al., 2012). In this case, each item would have to be unbound from its current position and re-bound to a new position, one back in the sequence.

6.5 Conclusion

The ability to update the contents of working memory is a critical element in the armory of skills needed for flexible cognition. The work presented in this thesis provides rigorous experimental evidence for what has to date been a limited, descriptive conceptualisation of working memory updating. A clear conclusion is that updating the serial order of items held in working memory imposes a high demand on general cognitive resources. At least in the temporal parameters of the standard paradigms used in the current research, updating-related demands appeared about one second from the onset of the update item and could not be shifted to accommodate faster paced tasks. This time course was reliably observed over multiple experiments using converging operations and generalised across two exemplar updating tasks. These findings provide the foundations critical for hypothesis-driven empirical research needed to identify the detailed nature of the multiple cognitive processes likely involved in updating.

Bibliography

Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology*, *76*(1), 129–140. https://doi.org/10.1037/h0025290

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology*, *3*, 301. https://doi.org/10.3389/FPSYG.2012.00301

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology*, *67*(1), 21–38. https://doi.org/10.1006/JECP.1997.2397

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698–1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Alloway, T. P., Kerr, I., & Langheinrich, T. (2010). The effect of articulatory suppression and manual tapping on serial recall. *European Journal of Cognitive Psychology*, *22*(2), 297–305. https://doi.org/10.1080/09541440902793731

Anders, T. R., & Lillyquist, T. D. (1971). Retrieval time in forward and backward recall. *Psychonomic Science*, *22*(4), 205–206. https://doi.org/10.3758/BF03332570

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language and Communication Disorders*, *41*(6), 675–693. https://doi.org/10.1080/13682820500442602

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, *7*(1), 25–31. https://doi.org/10.1111/j.1467-9280.1996.tb00662.x

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and Aging*, *5*(3), 421–428. https://doi.org/10.1037/0882-7974.5.3.421

Bäckman, L., Waris, O., Johansson, J., Andersson, M., Rinne, J. O., Alakurtti, K., … Nyberg, L. (2017). Increased dopamine release after working-memory updating training: Neurochemical correlates of transfer. *Scientific Reports*, *7*(1), 7160.

https://doi.org/10.1038/s41598-017-07577-y

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*(4), 302–309. https://doi.org/10.1080/14640746608400047

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Ed.), *Neuropsychological impairments of short-term memory* (pp. 54–73). New York, NY, US: Cambridge University Press.

Baddeley, A. D. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13468–13472. https://doi.org/10.1073/pnas.93.24.13468

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. D. (2006). Working Memory: An Overview. *Working Memory and Education*, 1–31. https://doi.org/10.1016/B978-012554465-8/50003-X

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2010). Investigating the episodic buffer. *Psychologica Belgica*, *50*(3–4), 223–243. https://doi.org/10.5334/pb-50-3-4-223

Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's Disease. *Brain*, *114*(6), 2521–2542. https://doi.org/10.1093/brain/114.6.2521

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. *Psychology of Learning and Motivation*, *8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2019). From short-term store to multicomponent working memory: The role of the modal model. *Memory & Cognition*, *47*(4), 575–588. https://doi.org/10.3758/s13421-018-0878-5

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233–252. https://doi.org/10.1080/14640748408402157

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science*, *21*(6), 413–419. https://doi.org/10.1177/0963721412459513

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology*, *45*(2), 477–490. https://doi.org/10.1037/a0014615

Barrouillet, P., Portrat, S., Vergauwe, E., Diependaele, K., & Camos, V. (2011). Further evidence for temporal decay in working memory: Reply to Lewandowsky and Oberauer (2009). *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(5), 1302–1317. https://doi.org/10.1037/a0022933

Bopp, K. L., & Verhaeghen, P. (2018). Aging and n-Back Performance: A Meta-Analysis. *The Journals of Gerontology: Series B*. https://doi.org/10.1093/geronb/gby024

Bor, D., Cumming, N., Scott, C. E. L., & Owen, A. M. (2004). Prefrontal cortical involvement in verbal encoding strategies. *European Journal of Neuroscience*, *19*(12), 3365–3370. https://doi.org/10.1111/j.1460-9568.2004.03438.x

Botto, M., Basso, D., Ferrari, M., & Palladino, P. (2014). When working memory updating requires updating: Analysis of serial position in a running memory task. *Acta Psychologica*, *148*, 123–129. https://doi.org/10.1016/j.actpsy.2014.01.012

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*(2), 563–570. https://doi.org/10.3758/BRM.42.2.563

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*(10), 1691–1700. https://doi.org/10.1080/17470210600848402

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, *31*(4), 429–460. https://doi.org/10.1016/0749-596X(92)90022-P

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. https://doi.org/10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. https://doi.org/10.1016/j.jml.2006.08.005

Camos, V., & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: the executive and phonological loops. *Frontiers in Human Neuroscience*, *8*, 1–11. https://doi.org/10.3389/fnhum.2014.00900

Carlesimo, G. A., Perri, R., Turriziani, P., Tomaiuolo, F., & Caltagirone, C. (2001). Remembering what but not where: independence of spatial and visual working memory in the human brain. *Cortex*, *37*(4), 519–534. https://doi.org/10.1016/s0010-9452(08)70591-4

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults? *Experimental Psychology*, *54*(4), 311–320. https://doi.org/10.1027/1618-3169.54.4.311

Carretti, B., Cornoldi, C., De Beni, R., & Romano, M. (2005). Updating in working memory: a comparison of good and poor comprehenders. *Journal of Experimental Child Psychology*, *91*(1), 45–66. https://doi.org/10.1016/j.jecp.2005.01.005

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: A computational model of the n-back task. *Journal of Cognitive Neuroscience*, *23*(11), 3598–3619. https://doi.org/10.1162/jocn_a_00047

Chen, T., & Li, D. (2007). The roles of working memory updating and processing speed in mediating age-related differences in fluid intelligence. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *14*(6), 631–646. https://doi.org/10.1080/13825580600987660

Chiappe, P., Siegel, L. S., & Hasher, L. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition*, *28*(1), 8–17. https://doi.org/10.3758/BF03211570

Cohen, J. D., Forman, S. D., Braver, T. S., Casey, B. J., Servan-Schreiber, D., & Noll, D. C. (1994). Activation of the prefrontal cortex in a nonspatial working memory task with functional MRI. *Human Brain Mapping*, *1*(4), 293–304. https://doi.org/10.1002/hbm.460010407

Collette, F., Van der Linden, M., Laureys, S., Arigoni, F., Delfiore, G., Degueldre, C., … Salmon, E. (2007). Mapping the updating process: Common and specific brain activations across different versions of the running span task. *Cortex*, *43*(1), 146–158. https://doi.org/10.1016/S0010-9452(08)70452-0

Conrad, R. (1964). Acoustic Confusions in Immediate Memory. *British Journal of Psychology*, *55*(1), 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, *4*(3), 161–169. https://doi.org/10.1016/S0022-5371(65)80015-9

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*(4), 429–432. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

Cornoldi, C., Drusi, S., Tencati, C., Giofrè, D., & Mirandola, C. (2012). Problem solving and working memory updating difficulties in a group of poor comprehenders. *Journal of Cognitive Education and Psychology*, *11*(1), 39–44. https://doi.org/10.1891/1945-8959.11.1.39

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, *31*(5), 668–684. https://doi.org/10.1016/0749-596X(92)90034-U

Cowan, N. (1995). *Attention and memory : an integrated framework*. Oxford University

Press.

Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In Akira Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 62–101). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short term memory. A reconsideration of storage capacity. *Behavioral and Brain Sciences*, *24*(4), 87–186. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2005). *Working memory capacity*. New York: Routledge. https://doi.org/10.4324/9780203342398

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, *145*(8), 822–847. https://doi.org/10.1037/bul0000199

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Cowan, N., & Rachev, N. R. (2018). Merging with the path not taken: Wilhelm Wundt's work as a precursor to the embedded-processes approach to memory, attention, and consciousness. *Consciousness and Cognition*, *63*, 228–238. https://doi.org/10.1016/j.concog.2018.06.001

Craik, F. I. . M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*. https://doi.org/10.3758/BF03210660

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510–1512. https://doi.org/10.1126/science.1155466

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex*, *11*(4), 341–354. https://doi.org/10.1016/s0010-9452(75)80026-8

Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*(10), 1189–1199. https://doi.org/10.1016/S0028-3932(98)00159-6

Della Sala, S., & Logie, R. H. (2002). Neuropsychological impairments of visual and spatial working memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *The Handbook of Memory Disorders, 2nd Edition* (2nd ed., pp. 271–292). Wiley-Blackwell.

Deschuyteneer, M., Vandierendonck, A., & Muyllaert, I. (2006). Does solution of mental arithmetic problems such as 2 + 6 and 3 x 8 rely on the process of "Memory updating"? *Experimental Psychology*, *53*(3), 198–208. https://doi.org/10.1027/1618-3169.53.3.198

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245. https://doi.org/10.1080/17470210600926075

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language*, *74*, 77–90. https://doi.org/10.1016/j.jml.2013.09.003

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: an experimental decomposition and individual differences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(1), 170–189. https://doi.org/10.1037/a0017891

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, *74*, 1–15. https://doi.org/10.1016/j.jml.2014.03.006

Elosúa, M. R., & Ruiz, R. M. (2008). Absence of hardly pursued updating in a running memory task. *Psychological Research*, *72*(4), 451–460. https://doi.org/10.1007/s00426-007-0124-4

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current*

Directions in Psychological Science, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 972–992. https://doi.org/10.1037/0278-7393.18.5.972

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation*, *44*, 145–199. https://doi.org/10.1016/S0079-7421(03)44005-X

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual Differences in Working Memory Capacity and What They Tell Us About Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex. In Akira Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 102–134). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.007

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, *55*(4), 587–600. https://doi.org/10.1016/J.JML.2006.06.002

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*(1), 59–79. https://doi.org/10.3758/BF03196257

Farrell, S., Oberauer, K., Greaves, M., Pasiecznik, K., Lewandowsky, S., & Jarrold, C. (2016). A test of interference versus decay in working memory: Varying distraction within lists in a complex span task. *Journal of Memory and Language*, *90*, 66–87. https://doi.org/10.1016/J.JML.2016.03.010

Fiore, F., Borella, E., Mammarella, I. C., & de Beni, R. (2012). Age differences in verbal and visuo-spatial working memory updating: Evidence from analysis of serial position curves. *Memory*, *20*(1), 14–27. https://doi.org/10.1080/09658211.2011.628320

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective & Behavioral Neuroscience*, *1*(2), 137–160.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K.

(2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*(5), 774–782. https://doi.org/10.3758/BF03198412

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition*, *26*(2), 263–276. https://doi.org/10.3758/BF03201138

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, *105*, 19–42. https://doi.org/10.1016/J.JML.2018.10.003

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, *22*(8), 1019–1037. https://doi.org/10.1002/acp.1407

Gathercole, S. E., Woolgar, F., Kievit, R. A., Astle, D., Manly, T., & Holmes, J. (2016). How Common are WM Deficits in Children with Difficulties in Reading and Mathematics? *Journal of Applied Research in Memory and Cognition*, *5*(4), 384–394. https://doi.org/10.1016/J.JARMAC.2016.07.013

Gilchrist, A. L., & Cowan, N. (2011). Can the focus of attention accommodate multiple, separate items? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(6), 1484–1502. https://doi.org/10.1037/a0024352

Göthe, K., & Oberauer, K. (2008). The integration of familiarity and recollection information in short-term recognition: modeling speed-accuracy trade-off functions. *Psychological Research*, *72*(3), 289–303. https://doi.org/10.1007/s00426-007-0111-9

Gugerty, L. J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, *3*(1), 42–66. https://doi.org/10.1037/1076-898X.3.1.42

Hamilton, P., & Hockey, R. (1974). Active selection of items to be remembered: The role of timing. *Cognitive Psychology*, *6*(1), 61–83. https://doi.org/10.1016/0010-0285(74)90004-8

Hartley, A. A., Speer, N. K., Jonides, J., Reuter-Lorenz, P. A., & Smith, E. E. (2001). Is the dissociability of working memory systems for name identity, visual-object identity, and spatial location maintained in old age? *Neuropsychology*, *15*(1), 3–17. https://doi.org/10.1037/0894-4105.15.1.3

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193–225. https://doi.org/10.1016/S0079-7421(08)60041-9

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*(2), 73–137. https://doi.org/10.1006/COGP.1998.0685

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained Memory: Error Patterns Rule out Chaining Models of Immediate Serial Recall. *The Quarterly Journal of Experimental Psychology*, *49*(1), 80–115. https://doi.org/10.1080/713755612

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology*, *25*(1), 104–111. https://doi.org/10.1080/14640747308400328

Hockey, R., & Hamilton, P. (1977). The basis of the primacy effect: Some experiments with running memory. *Quarterly Journal of Experimental Psychology*, *29*(1), 49–63. https://doi.org/10.1080/00335557743000026

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Jahanshahi, M., Saleem, T., Ho, A. K., Fuller, R., & Dirnberger, G. (2008). A preliminary investigation of the running digit span as a test of working memory. *Behavioural Neurology*, *20*(1–2), 17–25. https://doi.org/10.3233/BEN-2008-0212

Jaroslawska, A. J., Gathercole, S. E., Allen, R. J., & Holmes, J. (2016). Following instructions from working memory: Why does action at encoding and recall help? *Memory & Cognition*, *44*(8), 1183–1191. https://doi.org/10.3758/s13421-016-0636-5

Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L., & Bunge, S. A. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in Psychology*, *5*, 218. https://doi.org/10.3389/fpsyg.2014.00218

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., &

Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, *9*(4), 462–475. https://doi.org/10.1162/jocn.1997.9.4.462

Juvina, I., & Taatgen, N. A. (2007). Modeling control strategies in the N-Back task. In *Eighth International Conference on Cognitive Modeling* (pp. 73–78). New York, NY: Psychology Press.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. https://doi.org/10.1037/0096-3445.130.2.169

Kane, M. J., & Conway, A. R. A. (2016). The invention of n-back : An extremely brief history. *The Winnower*, *3*, 2003–2005. https://doi.org/10.15200/winn.146722.26397

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kessler, Y. (2017). The role of working memory gating in task switching: A procedural version of the reference-back paradigm. *Frontiers in Psychology*, *8*, 2260. https://doi.org/10.3389/fpsyg.2017.02260

Kessler, Y., & Meiran, N. (2008). Two Dissociable Updating Processes in Working Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1339–1348. https://doi.org/10.1037/a0013078

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(3), 738–753. https://doi.org/10.1037/a0035545

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic Bulletin and Review*, *22*(6), 1770–1776. https://doi.org/10.3758/s13423-015-0853-0

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358. https://doi.org/10.1037/h0043688

Kiss, I., Pisio, C., Francois, A., & Schopflocher, D. (1998). Central executive function in working memory: event-related brain potential studies. *Cognitive Brain Research*, *6*(4), 235–247. https://doi.org/10.1016/S0926-6410(97)00035-9

Klauer, K. C., & Zhao, Z. (2004). Double Dissociations in Visual and Spatial Short-Term Memory. *Journal of Experimental Psychology: General*, *133*(3), 355–381. https://doi.org/10.1037/0096-3445.133.3.355

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Kwong See, S. T., & Ryan, E. B. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging*, *10*(3), 458–468. https://doi.org/10.1037/0882-7974.10.3.458

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Lashley, K. S. (1951). The problem of serial order in behavior. In L. . Jeffress (Ed.), *Cerebral mechanisms in behavior; the Hixon Symposium* (pp. 112–146). Oxford, England: Wiley.

Lewandowsky, S., & Farrell, S. (2008). Short-Term memory: New data and a model. *Psychology of Learning and Motivation*, *49*, 1–48. https://doi.org/10.1016/S0079-7421(08)00001-7

Lewandowsky, S., & Heit, E. (2006). Some targets for memory models. *Journal of Memory and Language*, *55*(4), 441–446. https://doi.org/10.1016/J.JML.2006.08.006

Lewandowsky, S., & Oberauer, K. (2009). No Evidence for Temporal Decay in Working Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1545–1551. https://doi.org/10.1037/a0017010

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, *122*(4), 674–699. https://doi.org/10.1037/a0039684

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences*, *1424*, 33–44. https://doi.org/10.1111/nyas.13714

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(2), 229–247. https://doi.org/10.1080/14640748608401596

Logie, R. H. (1995). *Essays in cognitive psychology: Visuo-spatial working memory.* New

York, NY: Psychology Press.

Logie, R. H. (2016). Retiring the central executive. *Quarterly Journal of Experimental Psychology*, *69*(10), 2093–2109. https://doi.org/10.1080/17470218.2015.1136657

Logie, R. H. (2019). Converging sources of evidence and theory integration in working memory: A commentary on Morey, Rhodes, and Cowan (2019). *Cortex*, *112*, 162–171. https://doi.org/10.1016/j.cortex.2019.01.030

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*(1), 55–74. https://doi.org/10.1016/0001-6918(90)90066-O

Luzzatti, C., Vecchi, T., Agazzi, D., Cesa-Bianchi, M., & Vergani, C. (1998). A Neurological Dissociation Between Preserved Visual and Impaired Spatial Processing in Mental Imagery. *Cortex*, *34*(3), 461–469. https://doi.org/10.1016/S0010-9452(08)70768-8

Mackworth, J. F. (1959). Paced memorizing in a continuous task. *Journal of Experimental Psychology*, *58*(3), 206–211. https://doi.org/10.1037/h0049090

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*(1), 10–17. https://doi.org/10.3758/BF03195736

Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1272–1299. https://doi.org/10.1037/0278-7393.25.5.1272

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. *Memory and Cognition*, *44*(7), 1014–1037. https://doi.org/10.3758/s13421-016-0616-9

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Miyake, A., & Shah, P. (Eds.). (1999). *Models of Working Memory*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Moore, M. E., & Ross, B. M. (1963). Context effects in running memory. *Psychological Reports*. https://doi.org/10.2466/pr0.1963.12.2.451

Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin*, *144*(8), 849–883. https://doi.org/10.1037/bul0000155

Morey, C. C., Rhodes, S., & Cowan, N. (2019). Sensory-motor integration and brain lesions: Progress toward explaining domain-specific phenomena within domain-general working memory. *Cortex*, *112*, 149–161. https://doi.org/10.1016/J.CORTEX.2018.11.030

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*(1), 111–121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory and Cognition*, *44*(6), 922–936. https://doi.org/10.3758/s13421-016-0608-9

Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*. https://doi.org/10.1037/h0031844

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405–408. https://doi.org/10.1037/0033-295X.83.5.405

Murdock, B. B. (1966). Effects of a subsidiary task on short-term memory. *British Journal of Psychology*, *56*, 4. https://doi.org/10.1111/j.2044-8295.1965.tb00983.x

Murty, V. P., Sambataro, F., Radulescu, E., Altamura, M., Iudicello, J., Zoltick, B., … Mattay, V. S. (2011). Selective updating of working memory content modulates meso-cortico-striatal activity. *NeuroImage*, *57*(3), 1264–1272. https://doi.org/10.1016/j.neuroimage.2011.05.006

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Dori, H. (1998). Effects of divided attention on encoding and retrieval processes in human memory: Further support for an asymmetry. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.24.5.1091

Naveh-Benjamin, M., Craik, F. I. M., Perretta, J. G., & Tonev, S. T. (2000). The Effects of Divided Attention on Encoding and Retrieval Processes: The Resiliency of Retrieval Processes. *The Quarterly Journal of Experimental Psychology Section A*, *53*(3), 609–625. https://doi.org/10.1080/713755914

Norman, D. A., & Shallice, T. (1986). Attention to Action. In *Consciousness and Self-Regulation* (pp. 1–18). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-0629-1_1

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992–1009. https://doi.org/10.1037/bul0000108

Norris, D. (2019). Even an activated long-term memory system still needs a separate short-term store: A reply to Cowan (2019). *Psychological Bulletin*, *145*(8), 848–853. https://doi.org/10.1037/bul0000204

Norris, D., Hall, J., & Gathercole, S. E. (2019). How do we perform backward serial recall? *Memory and Cognition*, *47*(3), 519–543. https://doi.org/10.3758/s13421-018-0889-2

Oberauer, K. (2002). Access to Information in Working Memory: Exploring the Focus of Attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(3), 411–421. https://doi.org/10.1037//0278-7393.28.3.411

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*(3), 368–387. https://doi.org/10.1037/0096-3445.134.3.368

Oberauer, K. (2009). Chapter 2 Design for a Working Memory. *Psychology of Learning and Motivation - Advances in Research and Theory*, *51*(09), 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2013). The focus of attention in working memory—from metaphors to mechanisms. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00673

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences*, *1424*, 239–255. https://doi.org/10.1111/nyas.13603

Oberauer, K. (2019). Working memory and attention – A conceptual analysis and review. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.58

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in Immediate Serial Recall: Decay, Temporal Distinctiveness, or Interference? *Psychological Review*, *115*(3), 544–576. https://doi.org/10.1037/0033-295X.115.3.544

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73*(1), 15–30. https://doi.org/10.1016/j.jml.2014.02.003

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin and Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

Oberauer, K., & Vockenberg, K. (2009). Updating of working memory: lingering bindings. *Quarterly Journal of Experimental Psychology (2006)*, *62*(5), 967–987. https://doi.org/10.1080/17470210802372912

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

Öztekin, I., & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 383–397. https://doi.org/10.1037/a0018029

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition*, *29*(2), 344–354. https://doi.org/10.3758/bf03194929

Palladino, P., & Jarrold, C. (2008). Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology (2006)*, *61*(3), 392–399. https://doi.org/10.1080/17470210701664989

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processes in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences*, *15*(4), 257–269. https://doi.org/10.1016/J.LINDIF.2005.03.001

Pearson, D. G. (2006). Rehearsal processes in visuo-spatial working memory. *Cognitive Processing*, *7*(S1), 163–163. https://doi.org/10.1007/s10339-006-0123-1

Pearson, D. G., Ball, K., & Smith, D. T. (2014). Oculomotor preparation as a rehearsal mechanism in spatial working memory. *Cognition*, *132*(3), 416–428. https://doi.org/10.1016/j.cognition.2014.05.006

Pearson, D., & Sahraie, A. (2003). Oculomotor Control and the Maintenance of Spatially and Temporally Distributed Events in Visuo-Spatial Working Memory. *The Quarterly Journal of Experimental Psychology Section A*, *56*(7), 1089–1111. https://doi.org/10.1080/02724980343000044

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of*

Experimental Psychology, *57*(3), 137–146. https://doi.org/10.1037/h0046137

Portrat, S., Barrouillet, P., & Camos, V. (2008). Time-Related Decay or Interference-Based Forgetting in Working Memory? *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1561–1564. https://doi.org/10.1037/a0013356

Postle, B. R. (2003). Context in verbal short-term memory. *Memory & Cognition*, *31*(8), 1198–1207. https://doi.org/10.3758/BF03195803

Postle, B. R., Berger, J. S., Goldstein, J. H., Curtis, C. E., & D'Esposito, M. (2001). Behavioral and neurophysiological correlates of episodic coding, proactive interference, and list length effects in a running span verbal working memory task. *Cognitive, Affective & Behavioral Neuroscience*, *1*(1), 10–21.

Rac-Lubashevsky, R., & Kessler, Y. (2016). Decomposing the n-back task: An individual differences study using the reference-back paradigm. *Neuropsychologia*, *90*, 190–199. https://doi.org/10.1016/j.neuropsychologia.2016.07.013

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. https://doi.org/10.1037/a0029082

Rondeel, E. W. M., van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience*, *9*, 494. https://doi.org/10.3389/fnhum.2015.00494

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1019–1029. https://doi.org/10.1037/a0026976

Ross, B. M. (1966a). Serial-order effects in two-channel running memory. *Perceptual and Motor Skills*, *23*(3), 1099–1107. https://doi.org/10.2466/pms.1966.23.3f.1099

Ross, B. M. (1966b). Serial order as a unique source of error in running memory. *Perceptual and Motor Skills*, *23*(1), 195–209. https://doi.org/10.2466/pms.1966.23.1.195

Ruiz, R. M., & Elosúa, M. R. (2013). Conditions for positive and negative recencies in running memory-span recognition. *Acta Psychologica*, *144*(2), 213–223. https://doi.org/10.1016/j.actpsy.2013.07.006

Ruiz, R. M., Elosúa, M. R., & Lechuga, M. T. (2005). Old-fashioned responses in an updating memory task. *Quarterly Journal of Experimental Psychology Section A:*

Human Experimental Psychology, *58*(5), 887–908. https://doi.org/10.1080/02724980443000395

Salthouse, T. A. (2014). Relations between running memory and fluid intelligence. *Intelligence*, *43*(1), 1–7. https://doi.org/10.1016/j.intell.2013.12.002

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2017). Comparison of the Working Memory Load in N-Back and Working Memory Span Tasks by Means of EEG Frequency Band Power and P300 Amplitude. *Frontiers in Human Neuroscience*, *11*, 6. https://doi.org/10.3389/fnhum.2017.00006

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex Span Versus Updating Tasks of Working Memory: The Gap Is Not That Deep. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(4), 1089–1096. https://doi.org/10.1037/a0015730

Schmiedek, F., Li, S.-C., & Lindenberger, U. (2009). Interference and facilitation in spatial working memory: age-associated differences in lure effects in the n-back paradigm. *Psychology and Aging*, *24*(1), 203–210. https://doi.org/10.1037/a0014685

Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppe, R. A. (1996). PET Evidence for an Amodal Verbal Working Memory System. *NeuroImage*, *3*(2), 79–88. https://doi.org/10.1006/nimg.1996.0009

Shallice, T., & Warrington, E. K. (1970). Independent Functioning of Verbal Memory Stores: A Neuropsychological Study. *Quarterly Journal of Experimental Psychology*, *22*(2), 261–273. https://doi.org/10.1080/00335557043000203

Sheslow, D., & Adams, W. (1990). *Wide Range Assessment of Memory and Learning–Second Edition (WRAML-2)*. Jastak Associates. Retrieved from https://www.proedinc.com/Products/10501/wide-range-assessment-of-memory-and-learningsecond-edition-wraml2.aspx

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambric, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" Programs Work? *Psychological Science in the Public Interest*, *17*(3), 103–186.

Singh, K. A., Gignac, G. E., Brydges, C. R., & Ecker, U. K. H. (2018). Working memory capacity mediates the relationship between removal and fluid intelligence. *Journal of Memory and Language*, *101*, 18–36. https://doi.org/10.1016/j.jml.2018.03.002

Soveri, A., Karlsson, E. P. A., Waris, O., Gronholm-Nyman, P., & Laine, M. (2017). Pattern of Near Transfer Effects Following Working Memory Training With a Dual N-Back Task. *Experimental Psychology*, *64*(4), 240–252. https://doi.org/10.1027/1618-3169/a000370

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203–219. https://doi.org/10.1080/01443410903509259

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288. https://doi.org/10.1016/S0160-2896(01)00100-3

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of Interference During Working Memory Updating. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 137–151. https://doi.org/10.1037/a0020365

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., … Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology*, *10*(1), 1–10. https://doi.org/10.1186/1471-2288-10-1

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). Revisiting the attentional demands of rehearsal in working-memory tasks. *Journal of Memory and Language*, *105*, 1–18. https://doi.org/10.1016/j.jml.2018.10.005

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and Backward Recall: Different response time patterns, same retention order. *Psychological Science*, *14*(2), 169–174. https://doi.org/10.1111/1467-9280.01437

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A Reevaluation of Working Memory Capacity in Children. *Journal of Memory and Language*, *39*(2), 195–217. https://doi.org/10.1006/jmla.1998.2574

Turley-Ames, K. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*(4), 446–468. https://doi.org/10.1016/S0749-596X(03)00095-0

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Van Zandt, T., & Townsend, J. T. (2014). Designs for and Analyses of Response Time Experiments. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology, Vol. 1* (pp. 260–285). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199934874.013.0014

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, *95*(1), 57–79. https://doi.org/10.1348/000712604322779460

Verhaeghen, P., & Basak, C. (2005). Ageing and switching of the focus of attention in working memory: Results from a modified N-Back task. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(1), 134–154. https://doi.org/10.1080/02724980443000241

Vicari, S., Bellucci, S., & Carlesimo, G. A. (2003). Visual and spatial working memory dissociation: evidence from Williams syndrome. *Developmental Medicine and Child Neurology*, *45*(4), 269–273. https://doi.org/10.1017/s0012162203000513

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998–1003. https://doi.org/10.1111/1469-8986.3860998

Weems, S. A., Winder, R. K., Bunting, M., & Reggia, J. A. (2009). Running memory span: A comparison of behavioral capacity limits with those of an attractor neural network. *Cognitive Systems Research*, *10*(2), 161–171. https://doi.org/10.1016/j.cogsys.2008.09.001

Welford, A. T. (1958). *Ageing and human skill.* Glasgow:Oxford University Press.

Weschler, D. (1997). *Weschler Adult Intelligence Scale-Third Edition: Technical Manual*. San Antonio, TX: The Psychological Corporation. Retrieved from https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/-Wechsler-Adult-Intelligence-Scale-%7C-Third-Edition/p/100000243.html

Yntema, D. B. (1963). Keeping Track of Several Things at Once. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *5*(1), 7–17. https://doi.org/10.1177/001872086300500102

Yntema, D. B., & Mueser, G. E. (1962). Keeping track of variables that have few or many states. *Journal of Experimental Psychology*, *63*(4), 391–395. https://doi.org/10.1037/h0045706

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517.

https://doi.org/10.1006/JMLA.2002.2864

Zhao, X., Xu, Y., Fu, J., & Maes, J. H. R. (2018). Are training and transfer effects of working memory updating training modulated by achievement motivation? *Memory & Cognition*, *46*(3), 398–409. https://doi.org/10.3758/s13421-017-0773-5

Appendix A – Ethics approval letter

Karen Douglas
Secretary

Professor S Gathercole
MRC Cognition and Brain Sciences Unit
15 Chaucer Road
Cambridge
CB2 7EF

UNIVERSITY OF CAMBRIDGE

CAMBRIDGE
PSYCHOLOGY RESEARCH
ETHICS COMMITTEE

24 October 2016

Application No: PRE.2016.066

Dear Professor Gathercole

Assessing cognitive and neural processes of working memory updating

The Cambridge Psychology Research Ethics Committee has given ethical approval to your research project: "Assessing cognitive and neural processes of working memory updating", as set out in your application dated 4 August 2016.

The Committee attaches certain standard conditions to all ethical approvals. These are:

(a) that if the staff conducting the research should change, any new staff should read the application submitted to the Committee for ethical approval and this letter (and any subsequent letter concerning this application for ethical approval);
(b) that if the procedures used in the research project should change or the project itself should be changed, you should consider whether it is necessary to submit a further application for any modified or additional procedures to be approved;
(c) that if the employment or departmental affiliation of the staff should change, you should notify us of that fact.

Members of the Committee also ask that you inform them should you encounter any unexpected ethical issues.

If you would let us know that you are able to accept these conditions, we will record that you have been given ethical approval.

Please note that there have been changes to the procedures regarding amendments. Full details are given on the REC website.

Yours sincerely

K S Douglas

K S Douglas

cc Shraddha Kaur

Appendix B – Pilot studies for Experiment 1

Two pilot studies were conducted prior to Experiment 1 with the aim of establishing comparable the difficulty level across different working memory (WM) tasks by varying the WM load (denoted as *n*) such that average performance levels were comparable across tasks and fall within the target range of 70-85% accuracy.

Pilot study 1

Three groups of participants completed running span (*n* = 4 items), modified span (*n* = 6 items) and simple span (*n* = 8 items) respectively. Also, all participants completed a choice reaction time (RT) task and a dual-task in which both the memory task and the RT task were administered concurrently. A detailed description of the design, rationale, and task structures is provided in Section 2.3, and the recall accuracy and RT data from Pilot 1 are summarised in Table B1.

The observed data showed substantial variability in participant performance across tasks. While single-task performance in running span (= 80%) was within the target range, it was relatively higher in modified span (= 90%) and lower in simple span (= 50%).

The first pilot study also provided the opportunity to trial the utility of the divided attention method to generate an index of cognitive demand in the WM tasks. The RT data were analysed at both the task level as well as the trial level. The dual-task cost in choice-RTs was found to be higher when participants concurrently performed running span (= 126 ms) and modified span (= 166 ms) than simple span (= 71 ms). RT varied across the list as a function of serial position, with a steeper rise in running span than simple span (Figure B1). The periodic RT function in modified span suggested the demands of re-encoding new sequences intermittently with each target sequence within the longer modified span list treated similarly as the single list in simple span. These data indicated that the dual-task CRT paradigm did indeed provide a sensitive index to changes in cognitive load across the time course of the three WM tasks.

Table B1 – Data from pilot studies conducted before Experiment 1

			Recall accuracy		Reaction time (ms)	
	Group	N	Single	Dual	Single	Dual
Pilot 1	Running span *n* = 4 items	3	0.83 (0.10)	0.77 (0.11)	538 (84)	664 (88)
	Modified span *n* = 6 items	4	0.90 (0.02)	0.77 (0.02)	481 (76)	647 (183)
	Simple span *n* = 8 items	4	0.50 (0.08)	0.37 (0.05)	464 (57)	535 (65)
Pilot 2	Modified span *n* = 7 items	5	0.73 (0.14)	.	.	.
	Simple span *n* = 7 items	4	0.82 (0.19)	.	.	.

Note: Recall scored as a proportion of items recalled in the correct target position. *n* indicates the number of items participants were asked to recall in the respective WM tasks. N denotes the number of participants recruited per group.

Pilot study 2

A second pilot study was conducted to establish the size of target sets for modified span and simple span that were closer to the optimal level of 80%. Based on these data (Table B1), the WM loads of seven for both modified span and simple span and four for running span were chosen for Experiment 1.

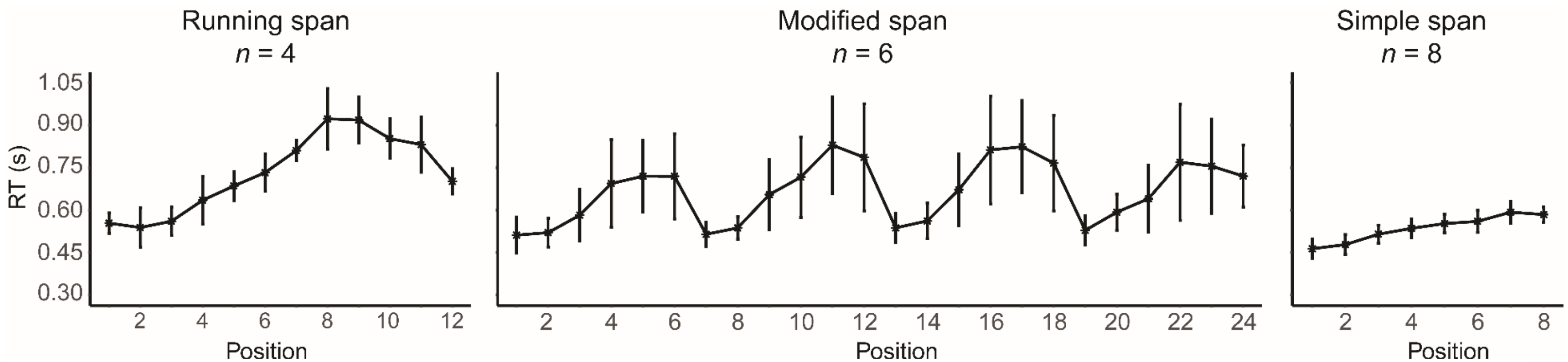

Figure B1: The reaction time (RT) data from the first pilot for the experiment reported in Chapter 2. The RTs from the concurrent CRT task are presented as a function of the serial position in the respective memory tasks completed by three groups. n denotes the number of itemsrequired for recall in each task. NB: the WM load (*n*) in this pilot for modified span and simple span were different from that used in the main study.

Appendix C – Pilot studies for Experiment 2

Two pilot studies were conducted prior to Experiment 2 to determine the rate of presentation most appropriate to elicit the difference, if any, between passive and active strategies.

Pilot study 1

The aim of Pilot 1 was to compare behaviour in running span and simple span at two presentation rates. Two groups with five participants each were recruited for this pilot study such that they completed different memory tasks. The divided attention approach, as described in Chapter 2, was also employed here in which the respective memory task was concurrently applied with the choice reaction time (CRT) task. Thus, each participant completed the respective memory task at two rates (500 ms per item and 800 ms per item) and two attentional loads (single and dual). The RT measure from the concurrent CRT task was used to index cognitive demands imposed by each variant of the two memory tasks.

It was observed that recall accuracy was higher in running span than simple span in both rate conditions (Table C1). Anecdotal reports by participants in the running span group indicated that they adopted passive strategies to perform the task rather than actively updating the target set as the list grew longer. This was likely due to the relatively fast presentation rates employed in this pilot study. Conversely, participants completing simple span reported employing a variety of strategies with a mix of active and passive information processing. The RTs from the concurrent task increased over the course of a trial in both memory tasks, and longer RTs were associated with the slower rate condition, especially in running span (Figure C1). The interval between items was further divided into two bins in both rate conditions. In the condition presenting items every 500 ms, the duration of the bins was 250 ms; in the condition presenting items every 800 ms, the bins were 400 ms in duration. The concurrent RTs were then examined as a function of these bins across early and late items in both rate and task conditions (Figure C2). Early items included those in positions one to four; late items included positions five onwards. There appeared to be little effect of either bin or rate on concurrent RTs during the early and late items in simple span trials. The only exception was that RTs appeared to speed up during late positions in the faster rate condition, but it was not clear why this may be the case. In running span, there appeared to be a general difference between early and late items, particularly at the slower rate condition of 800 ms per item, which echoed the data presented in Figure C1. There was little indication of any difference between bins for either early or late positions across the two rates in running span (Figure C2).

Table C1

Mean (SD) performance data from Pilot 1 of Experiment 2

	Task	Rate	N	Recall accuracy		Reaction time (ms)	
				Single	Dual	Single	Dual
Pilot 1	Running span *n* = 4 items	500 ms/item	5	.90 (.02)	.84 (.09)	434 (33)	399 (22)
		800 ms/item	5	.89 (.04)	.85 (.04)	459 (26)	429 (8)
	Simple span *n* = 7 items	500 ms/item	5	.52 (.09)	.40 (.08)	515 (29)	464 (24)
		800 ms/item	5	.60 (.07)	.43 (.10)	474 (26)	469 (40)
Pilot 2	Running span *n* = 4 items	400 ms/item	4	.76 (.03)	.73 (.03)		424 (22)
		800 ms/item	4	.81 (.02)	.76 (.05)	465(17)[1]	493 (30)
		1200 ms/item	4	.86 (.05)	.69 (.09)		550 (67)

Note: Recall scored as a proportion of items recalled in the correct target position. *n* indicates the number of items participants were asked to recall in the respective WM tasks. N denotes the number of participants recruited per group.

[1] In Pilot 2, the single CRT task was administered only once per participant and thus only presented here as a group mean across rate conditions. In Experiment 2, the single CRT task was repeated for each session in which the different rate conditions were administered.

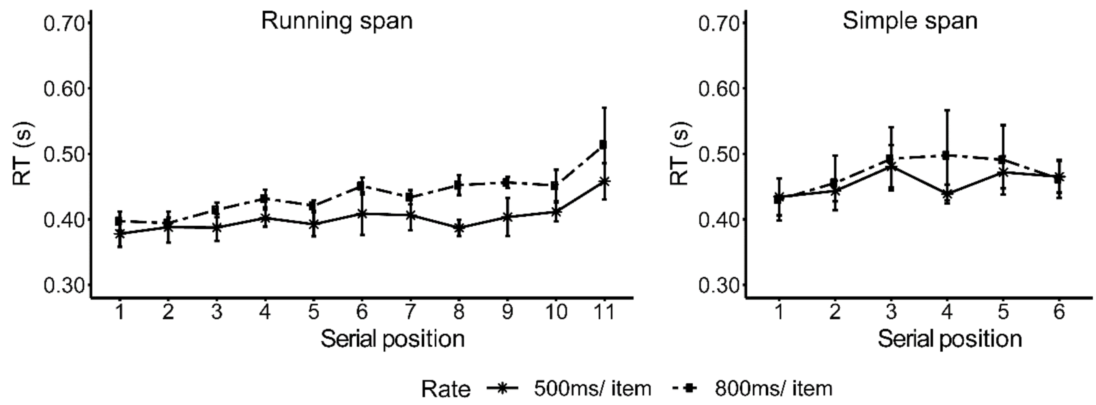

Figure C1 Mean concurrent RTs during the running span (left) and simple span (right) tasks, for the 500 ms/item (solid line) and 800 ms/item (dashed line) rate conditons in Pilot 1. Error bars represent the standard error of the mean.

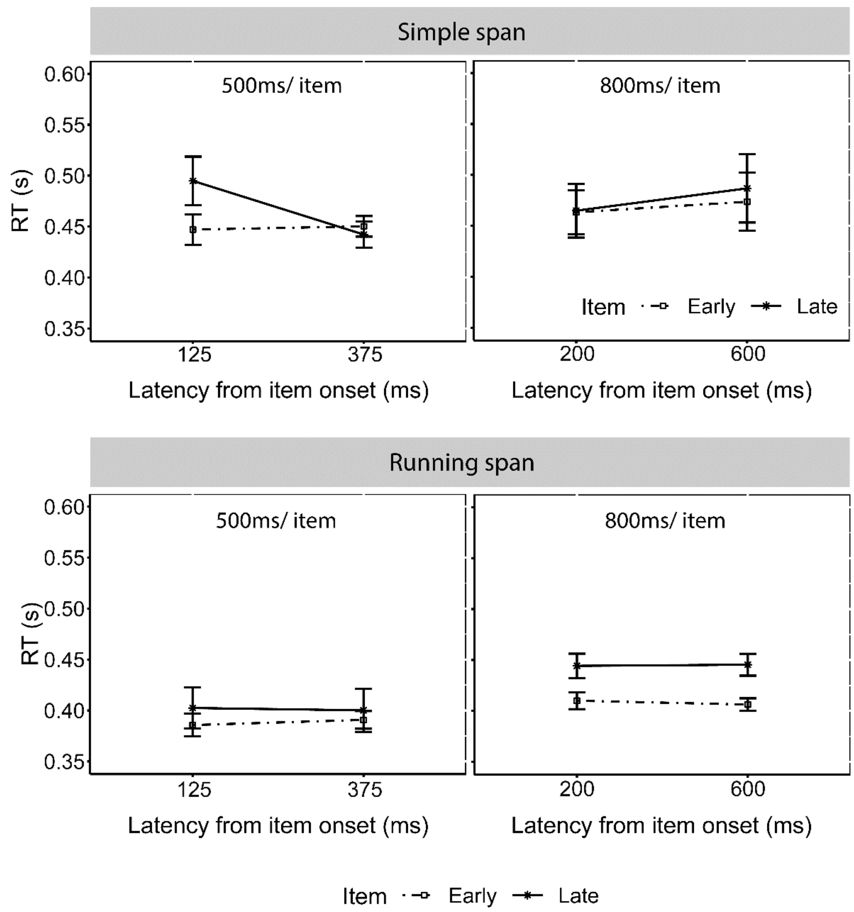

Figure C2 Mean concurrent RTs at the interval level across early positions (dashed line) and late positions (solid line) for simple span (top) and running span (bottom), with the 500 ms/item (left) and 800 ms/item (right) in Pilot 1. Error bars represent the standard error of the mean.

Pilot study 2

The data from the first pilot study suggested that simple span might not provide an appropriate comparison for running span, as the variation in presentation rate led to the adoption of different strategies in the two tasks. Further, it was found that the rates employed in Pilot 1 were unable to capture the expected shift from active updating to passive listening in running span. It was speculated that a larger difference between presentation rates might be required. Reflecting on the updating-related peak found 1000 ms following item onset in Experiment 1, the design was reconsidered to include a condition in which items were presented slower than an item per second.

This led to a second pilot study in which only running span was administered at three rate conditions (fast = 400 ms/item; medium = 800 ms/item; slow = 1600 ms/item). Four participants were recruited for this pilot study and each completed all three variants of the running span task, while also concurrently performing the CRT task. Recall accuracy improved as the presentation rate was slowed, with a substantial dual-task cost at the slow presentation rate (Table C1). It was observed that concurrent CRTs task increased with serial position in all rate conditions, but the slope appeared steeper in the slow and medium rate than the fast rate (Figure C3).

The interval between items in the medium and slow rate condition was divided into smaller bins of 400 ms each, and concurrent RTs were examined as a function of the bins, across the rate conditions and separately for early positions (items one to four) and late positions (items five onwards). For all three rates, the late positions in the running span task were associated with slower concurrent RTs compared with the early positions (Figure C4). It was also observed that there was some degree of variation across the item interval, but the data were likely too noisy to reveal any specific trends. Importantly, the data demonstrated a successful manipulation of resource demands in running span indexed by CRTs from the concurrent task by varying the rate of presentation. As such, this design was adopted for Experiment 2.

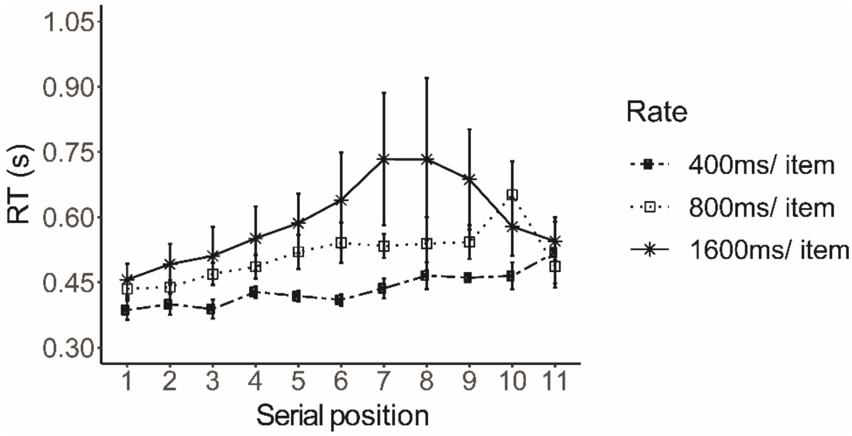

Figure C3 Mean concurrent RTs during running span with varying rates of item presentation in Pilot 2. The RTs are presented as a function of serial positions within a memory sequence. Error bars represent the standard error of the mean.

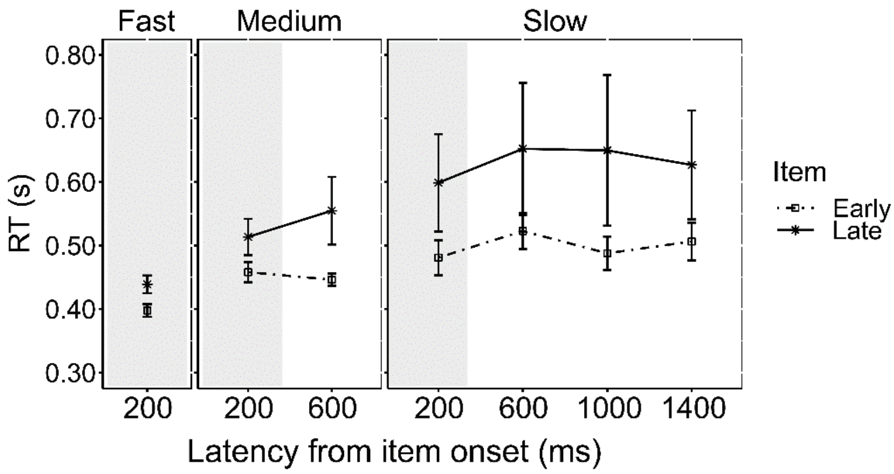

Figure C4 Mean concurrent RTs in Pilot 2 at the interval level across early positions (dashed line) and late positions (solid line) in the three rate conditions, fast (400 ms/item, left), medium (800 ms/item, middle), and slow (1600 ms/item, right). The first 400 ms represent the duration of item presentation (region shaded in grey), followed by a variable silent inter-item interval as per rate condition (unshaded region). Error bars represent the standard error of the mean.

Appendix D1 – Pilot studies for Experiment 3

Two pilot studies were conducted prior to Experiment 3. The main aim of this experiment was to find a reliable index of cognitive demand in running span that was derived from the memory task itself, rather than a concurrent task as in Experiment 1 and Experiment 2. Two pilot studies were conducted to explore the utility of a self-paced task in which participants could regulate the presentation of memory items. The presentation times could then be used as direct indices of ongoing demand.

Pilot study 1

The first pilot examined presentation times as well as recall accuracy in a self-paced running span and compared them with those derived from a self-paced version of simple span. A within-subject design was employed so each participant completed both running span and simple span tasks. In running span, participants had to retrieve the last four items from lists of unpredictable length; in simple span, the list length was fixed at seven and serial recall of all items was required. Both tasks were self-paced, i.e. participants pressed a key for the onset of each memory item, and these presentation times were recorded. The data, presented in Table D1, showed that cognitive processing during running span was more time-consuming than simple span. The serial position curves (see Figure D1) were similar to those recorded in Experiment 1 in this thesis, suggesting that participants perform similarly in both self-paced and experimenter-paced paradigms.

Further, the presentation time data in the running span condition were in line with expectations (see Figure D2). Participants appeared to favour speeded presentation for non-update items and slow presentation for update items, and this trend was similar across participants. The function in simple span was less stable and exhibited high variability across participants. It was possible that this variability in simple span presentation times reflected the variability in the strategies employed by participants. For instance, chunking a list of seven items as a sequence of 2-2-3 or 3-4 might result in very different profiles. The simple span task thus appeared to be vulnerable to individual differences and was deemed less suitable as a comparison task for running span.

Table D1 – Recall data from two pilot studies for Experiment 3

	Condition	N	Recall accuracy[1]		Presentation time (s)	
			Mean	SD	Mean	SD
Pilot 1	Running span (n = 4 items)	4	.88	.09	2.46	1.10
	Simple span (n = 7 items)	4	.80	.13	2.01	.51
Pilot 2	Active, running span (n = 4 items)	5	.81	.08	1.96	.73
	Passive, running span (n = 4 items)	5	.74	.12	.70	.29

[1] Recall accuracy was scored as the number of items recalled in the correct serial position

[2] Both pilots were designed as within-subject investigations.

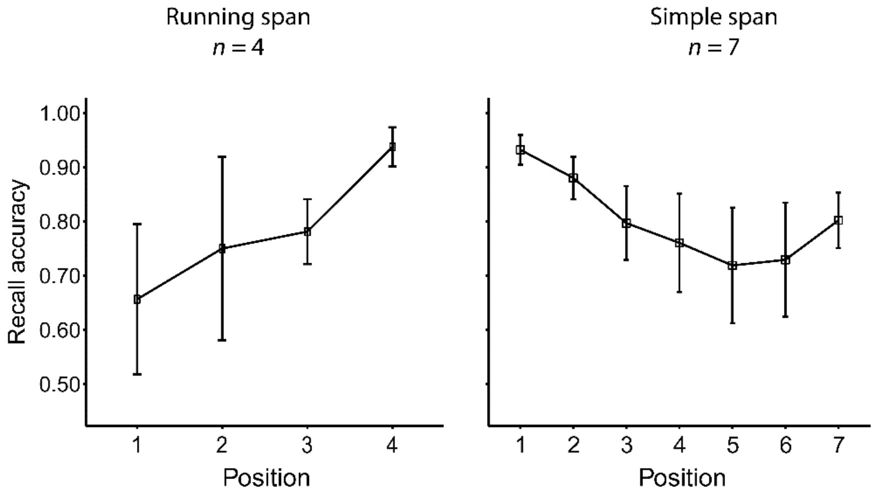

Figure D1 The recall accuracy at each output position in the running span (left) and simple span (right) using self-paced tasks from the first pilot of Experiment 3.

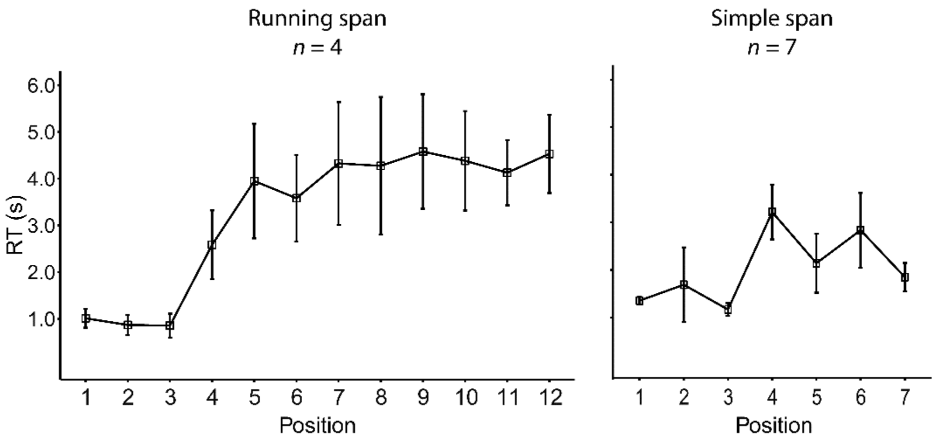

Figure D2 The presentation times at each input position in running span (left) and simple span (right) using self-paced tasks from the first pilot of Experiment 3.

Pilot study 2

A second pilot study was then conducted in which the presentation time functions associated with passive and active strategies in running span were considered. In this, a self-paced running span was administered and participants were instructed to either actively or passively perform the task. See Table D1 and Figure D3 for both presentation times and recall data.

The presentation time function associated with running span in the first pilot was replicated in the second pilot (active strategy condition), suggesting highly stable behavioural patterns. Further, the cognitive demands associated with the passive condition were low and position-invariant, suggesting that participants did not actively recruit resources to support memory performance during the passive task condition.

In addition to reliable data, the comparison of active versus passive modes in running span offered advantages in terms of task design. These conditions used the same running span task structure, thus presented the same variability in list length and required participants to recall four items in both conditions. In light of these advantages, the design from the second pilot was adopted in Experiment 3.

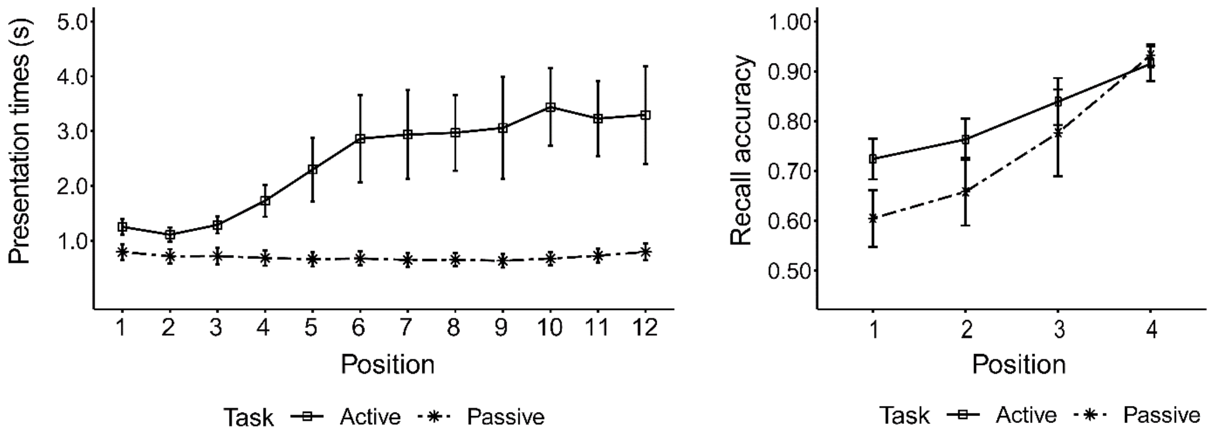

Figure D3 Data from the second pilot of Experiment 3 comparing presentation times (left) and recall (right) between the active and passive strategy conditions in a self-paced running span task.

Appendix D2 – Open strategy reports

The following are the open strategy reports collected from participants after they completed the running span task using an active or passive strategy. The original wording is preserved as much as possible, only adapted to form complete, legible sentences as participants often wrote incomplete thoughts. Very rarely, the experimenter also had to interpret a word or so as the written script was unclear. These changes are minimal and have been marked using square parentheses in the following tables. See Table D2 for the active strategy reports and Table D3 for the passive strategy reports.

Table D2.
Strategy reports following the active strategy condition

ID	Condition	Strategy report
578	Active	I was repeating the last 4 letters each time and pressed the keys much slower than I had [done] before [in the previous session] to make sure I was remembering. I also took more breaks.
579	Active	[I] pressed spacebar quickly when the first few letters were being presented because most lists consisted of more than 5 letters. [After that I] slowed down and tried to form a mental image of the letters subsequently. [I also] tried to associate letters with real-world places/organisations/ etc. e.g. EY
580	Active	I repeated the sound of the last 4 letters under my breath/mentally. I visualised the sequence of letters in a line. I recognised abbreviations/letter combinations.
581	Active	[For] the first four letters presented, I clicked through rapidly. This [rate of key-pressing] then slowed as more letters followed.
582	Active	At first, I tried just to repeat the last four letters after each new letter but was slow [at] recalling [them] so I tried dividing the

		letters into initials (MF, VK) and imagining them in block print above computer screen. This was still slow but better than the previous method. Eventually, I found [that] using [the] tone to be the most effective, repeating the previous four letters with an upward inflexion and new 4 with a downward inflexion. Each time a new letter came up I would say the previous 4 outbound going up and the new 4 going down.
583	Active	I did not begin with a strategy but after a few attempts, I began to separate the first of the four letters from the last three to aid my recall. In other words, I would repeat in my mind and also make the letter shapes with my mouth without vocalizing them in the form z-SGH with the stress on the last three letters. Upon hearing the next letter, I would add it to the final 3 normally first, ie. SGHP, and then convert it to the form s-GHP and repeat. Sometimes I would realise I had stopped consciously understanding the letters and was only repeating a mnemonic device which sometimes led me to confuse similar sounding letters. When I would notice this confusion I refocussed on both audible and conscious repetition.
584	Active	I repeated the sequence of letters in mind each time I heard a new letter.
585	Active	[I was] grouping letters by 1:3, so remembering the sequence as (a)(bcd) then as the list continued, shifting the group of 3 as the new beginning of list of 4: (bcd)(e) and then reconfiguring to (b)(cde) so the next letter added (cde)(f) --> (c)(def).
586	Active	I would listen to the first 4 letters and when one more letter came along, I would forget the first letter of the sequence. I would also say the sequence out loud for example - if the sequence was FBSH I would say that and when another letter came up, I would forget the F so it would be BSHR and again I would say this aloud.

587	Active	I tried a couple of different strategies: (1) first I tried listening alone, (2) then switched to repetition quite quickly, pausing after each letter given to repeat the sequence, and (3) when this proved too difficult, I looked at the keyboard and touched the letters being said. This allowed me to drop excess letters previous to the last four given and gave me something visual to remember the sequence by.
588	Active	When I was presented with the letters, I tried to recall in my head the last three letters and then add the new letter to the end and kept repeating this.
589	Active	When the letters were being presented, I made sentences with the letters (most short ones [Goats like rats; lions make quilts], but sometimes long ones [Go, Kazakstan, because yellow dogs xylophone]). At first, I tried to make each one different, in order to be distinct or memorable, but sometimes that backfired and I had trouble remembering anything. So I tried to make it less hard [always thought of kangaroos for K, etc] and repeat animals/verbs [this likes this, etc]. With the really long strands, I'd abandon the beginning phrase of the sentence and try to focus on the most recent "episode" of the sentence/story/weird narrative (snails press flat knots). In the first trial, I tried to make alphabet signs in American sign language as I thought that body memory would help but it confused me or it was too much to recall so I abandoned that practice. I closed my eyes to shut out visual stimulation and focus on the sentences.
590	Active	Much less stress (than previous passive condition) and was now happy to try and remember letters! Used the keyboard to memorise the last 4 letters for each task. Had great difficulty distinguishing similar sounding letters B, P, Q, U, etc.
591	Active	I repeated each letter in my head in the pattern it was presented. I repeated it numerous times. I also tried saying the letters out

		loud. I used my fingers to count the number of letters so that I always knew 4 letters. I tried only remembering the last 4 by forgetting the first letter when a new letter was spoken. I also tapped my fingers to a tune to remember the pattern.
592	Active	[I was] blocking letters in twos.
593	Active	It helped to say the letters through under my breath, moulding letters; I said the 4 letters to be remembered through once then the 3 I could carry forward if there was another letter following. I also, while I did this, pictured the letters in my head as strings of 4 and then the 3 carried forward.
594	Active	Grouped the letters into columns of 3 in my mind. Mainly ended up remembering the whole sequence.
595	Active	I had my eyes open, focussed on one particular object in the room at times, which worked well. I also matched the letters I was hearing with the letters on the keyboard. I avoided closing my eyes as this was distracting and I couldn't recall the letters as well. As the letters were read out to me, I counted 4 letters on my fingers so that I was able to remember which were the last 4 letters I heard.
596	Active	I started by remembering the first four and said them in my mind. As more letters were added, I would then take off the previous letter to keep the current four last letters in my mind and would repeat before listening for the next letter. I would repeat this to try and remember.
597	Active	[I] repeated the pattern aloud in my mind without hesitation and tried to answer quickly so I didn't forget.
598	Active	Yesterday I thought that was hard and remember deliberately was easy, but now I think remembering on purpose is hard!! As the letters were presented, I blocked them in groups of 4 and

		tried to remember each block until the next block was complete. Then I would forget about the first block and just keep the most recent one in mind. Still, if distracting thoughts popped into my mind then it would mess up my blocks and I'd have to pause or work harder to remember them. As time went on, all that concentrating felt quite exhausting!
599	Active	[I would] visualise letters being typed out, keep 'spotlight' on last 4 letters, plenty of mental repetition - helps remember further back letters.

Table D3.

Strategy reports following the passive strategy condition

ID	Condition	Strategy report
578	Passive	For the first round of tests, I was repeating letters in my head so I could remember them. However, after the reminder came up to listen passively I stopped doing this. I listened to each letter as they came up.
579	Passive	[I was] pressing spacebar as quickly as possible so I don't form a mental image of the letters.
580	Passive	I initially was keeping the last four letters in my head (without trying to do so) but I was deliberately trying to focus on the cross on the screen and ignore any understanding of the letters beyond the sound.
581	Passive	[I] tried to avoid focussing too much on the letters, thereby preventing rehearsal in my head, via staring at the desk or wall.

582	Passive	The closest I got to [developing a] strategy was pressing the spacebar after each letter so the gap between them was pretty regular but that was more or less coincidental. I would hear the letter again immediately after it was spoken.
583	Passive	I had to actively stop myself from rehearsing the list of letters. As I heard them I would automatically begin to rehearse them in order in a manner that would aid recall.
584	Passive	I tried to listen to the letters as groups of 4. I paid attention more when I thought it was coming to the end of the list in order to try and remember the last 4 letters.
585	Passive	[I] stared at fixation cross, went at faster speed so I could not dwell on memorising the letters. Some letter combinations were impossible not to group since I have personal associations (e.g. JT = Justin Timberlake).
586	Passive	I tried my hardest to listen passively so I would look around the room and sing songs in my head. Occasionally, however, a few of the letters would form an acronym I was familiar with such as BK (burger king) and so this would stick in my mind. That being said, these acronyms would appear at the beginning or middle of the sequence so wouldn't help me recall.
587	Passive	All I did was try and keep my eyes moving around the room and distract myself while listening.
588	Passive	I tried to concentrate on other parts of the room whilst passively listening to the letters.
589	Passive	In order to remain a “passive listener”, I tried to stay very aware of my surroundings / body awareness so that I was not honed in too much on memorising / chunking that I wanted to be doing, i.e. I looked around the room, rubbed my knees and visualised the letters as they were said. I thought of the letters in capitals (not lower case) and in different colours.

590	Passive	I [pressed keys] relatively [quickly] for the first four letters, then [used] a pause to remember them. Initially, [I] could not help but try to memorise letters as they came along, and [then I] began grouping them. At points, I realised that this was not how I should have been performing the task so then just tried to remove the letters from my mind, then on the next key press, I would have to recall. Found it very difficult to not pay attention to the letters. I didn't know where to look or to focus so my mind would constantly drift as more letters were presented. [I] did find [that] at points I was trying to make patterns using the keyboard to remember. [There were] lots of delays on spacebar press as I became more frustrated.
591	Passive	I found it difficult to passively listen to the letters. I had to try and not use my fingers to count the letters. I also had to stop myself from saying the letters in my head in the order. I tried repeating each letter after I heard it. I also found going through the list faster helped me to listen passively as listening to them quickly did not give me the time to think deeply about what I was listening to.
592	Passive	[I was] trying to keep my mind 'blank'. [This] was almost harder than when grouping [in other condition].
593	Passive	To listen passively I tried to look around the room to stop myself from remembering or rehearsing letters.
594	Passive	[I] just thought more about the spacebar. [I] allowed myself to daydream a bit. [I] got the letters fast rather than waiting.
595	Passive	I tried not to focus on the letters as much as possible and instead looked at the objects around the room, sung songs in my head, or thought about things unrelated to the task. I also laid back in the chair and occasionally closed my eyes.

596	Passive	I tried hard to not memorise the letters presented. I did this by pressing the space bar so quickly they said quickly. This way I was trying to listen passively and not using a strategy to remember them.
597	Passive	No strategy [was] adopted as [I] was just listening to the letters. Tried to go fast to not allow me to memorise.
598	Passive	It's so hard not to group the letters together!! That was my normal instinct - to sort of the block the letters into groups of 4. I had to try and resist doing that by just letting the letters wash over and maybe paying attention to each one as it appeared but not keeping it in mind longer. But in doing this, I often felt like I had no recollection of the letters or order so spontaneously I would just start grouping them again, so I would have to resist temptation again! It was also harder the longer the lists, short lists of letters were easier to just sort of passively absorb.
599	Passive	Repeat the last letter to try not to group

Appendix E1 – Pilot studies for Experiment 4

Two pilot studies were conducted before Experiment 4 with the overall aim of extending the investigation from running span to another task involving serial updating, namely the *n*-back task.

Pilot study 1

The aim of the first pilot study was to determine the feasibility of inducing different strategies to complete the *n*-back task. This pilot study adopted the experimental design used in Experiment 3. Participants were instructed to use either an active or a passive strategy to perform the *n*-back task. The *n* was fixed at four, comparable with the running span investigations conducted previously. Participants pressed the left arrow key if the most recent item matched the item presented four positions ago and they pressed the right arrow key if the items were different. A self-paced procedure was used, such that participants used keypresses to regulate the presentation time of each item. The keypress used to indicate item comparison decisions also triggered the onset of the next item.

The presentation times measured in this way during both the active and passive strategy condition were compared (Table E1). Participants preferred slower item presentation when they use an active than passive strategy in the *4*-back task (mean difference = 170 ms). This overall difference between the strategy conditions however did not show a consistent pattern across the trial (Figure E1). This was inconsistent with results from Experiment 3 that showed that an effect of strategy on presentation times emerged at the fourth (= *n*th) item in the trial.

A number of features of the design of this pilot study suggested it was unsuitable for a full experiment. First, *n*-back was not conducive to the adoption of a passive strategy. Participants found it challenging, if not impossible, to employ a passive approach in *n*-back while also continuously having to compare the most recent item with an older one and judge if the items matched. They reported being unable to simply listen passively to incoming items without engaging active maintenance or updating strategies when another aspect of the task required active recognition judgements. Second, it was not possible to disentangle the time taken for the *n*-back recognition judgement and that taken to update working memory. Both components of the task were indicated by the same keypress, which obscured the unique contribution of each.

Table E1 – Data from pilot studies conducted prior to Experiment 4

	N-back condition	N	*n*	Reaction time (s) Single	Reaction time (s) Dual
Pilot 1	Active strategy	3	4	1.09 (1.02)	.
	Passive strategy	2	4	.92 (.65)	.
Pilot 2	*1*-back	1	4	.47 (.08)	.49 (.11)
	4-back	4	1	.46 (.08)	.63 (.19)

Note: n indicates the number of positions back in the list participants were asked to compare with the current item. N denotes the number of participants recruited per group.

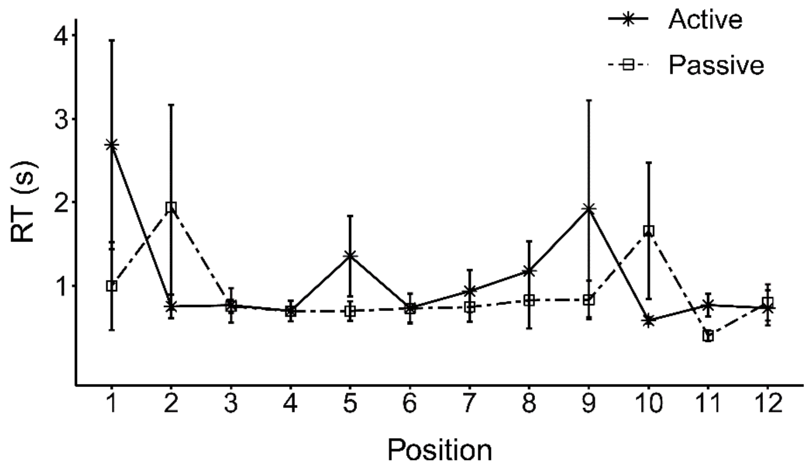

Figure E1 . Inter-item intervals across list positions during the *4*-back task during both active (solid line) and passive strategy (broken line). Note that the data are averaged across all list lengths, thus later positions contribute fewer data points. Error bars represent standard error of the mean.

Pilot study 2

A second pilot study tested the feasibility of applying the divided attention method used in Experiments 1 and 2 to the *n*-back task. *1*-back and *4*-back conditions were studied using the same task and trial structures as in previous running span experiments. Each *n*-back condition was administered independently and then simultaneously with a choice reaction time task (CRT). The concurrent CRTs were used to index the demands imposed by *n*-back and these were compared across the *1*-back and *4*-back conditions (Table E1). Results showed that the CRTs in the single load condition were similar in the two *n*-back conditions. There was an increase in CRTs under dual load condition in the *4*-back but not *1*-back task, indicating that the condition with a higher *n* demanded greater resources.

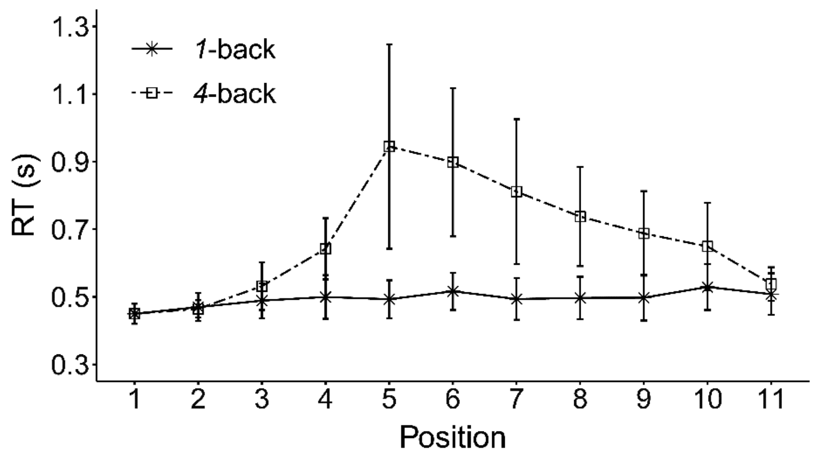

Figure E2 Concurrent CRTs across list positions for *1*-back (solid line) and *4*-back tasks (broken line). Note that the data are averaged across all list lengths, thus later positions contribute fewer data points. RTs associated with the final position across lists are not displayed here, see text for data exclusion. Error bars represent standard error of the mean.

The CRTs in the concurrent task over the course the trials in both *n*-back conditions are summarised in Figure E2. CRTs increased with serial position in the more demanding *4*-back condition and decreased towards the end of the sequence. This is in contrast with the CRT function during concurrent running span, in which the CRTs did not decrease after the fifth item in the trial. This could be due to the differences in the composition of the running span and *n*-back tasks. In the running span tasks administered across Experiments 1-3, sequences contained up to 12 items and item repetition within the same sequence was minimised so that most items were unique. These conditions are different from the typical *n*-back sequences that are relatively lengthy with 20-40 items (Awh et al., 1996; Hartley et al., 2001; Jaeggi et al., 2010; Kane et al., 2007; Owen et al., 2005; Szmalec et al., 2011). Long sequences enable the items to be regularly repeated (providing enough targets) while allowing a sufficient number of unique items to decrease target predictability. In this way, trials are typically structured so that 20-25% of the items are *n*-back targets. The use of list lengths typical for running span in the *n*-back conditions in this pilot may therefore not be appropriate.

In summary, two pilot studies revealed that a divided attention approach was more suitable than the self-paced paradigm to examine the serial-updating process in *n*-back tasks. It was also found that the *n*-back task required trial structures (in terms of the number of repeated and unique items) different from those used in the studies of running span. Therefore, Experiment 4 administered *n*-back sequences that contained 20+*n* items with 6 targets in each sequence. Two *n*-back conditions were used in Experiment 4, *1*-back and *4*-back, and both were administered under divided attention conditions to enable the measurement of resource demands over the course of each condition.

Appendix E2 – Supplementary data analysis in Experiment 4

Paired-sample t-tests of each successive pair of positions in the *4*-back task.

Positions	*Mean difference (ms)*	*t*	*p*	*Cohen's d*
1 vs 2	13	2.24	**.03**	.42
2 vs 3	32	3.60	***.001***	.78
3 vs 4	42	5.18	***<.001***	1.45
4 vs 5	21	1.66	.11	.31
5 vs 6	- 1	.152	.88	- .02
6 vs 7	2	.31	.759	.06
7 vs 8	- 3	.39	.70	- .09
8 vs 9	3	.45	.66	.10
9 vs 10	<1	.01	> .99	<.01
10 vs 11	- 2	.27	.79	- .05
11 vs 12	9	1.09	.28	.21
12 vs 13	- 7	1.02	.32	- .21
13 vs 14	- 11	1.51	.14	- .29
14 vs 15	8	1.10	.28	.21
15 vs 16	18	1.47	.15	.32
16 vs 17	- 22	1.50	.14	- .30
17 vs 18	- 6	.53	.59	- .02
18 vs 19	8	1.76	.09	.07
19 vs 20	8	.86	.40	.16
20 vs 21	- 11	1.55	.13	- .29
21 vs 22	9	1.16	.26	.26
22 vs 23	- 22	1.76	.09	- .44

Note: Bold text denote significant effects at the $p < .05$ level, bold italicized text indicate significance effects after adjusting for multiple comparisons using the Bonferroni correction method. Negative mean difference and Cohen's d values reflect a decrease in RT between consecutive position.